New Physics at the LHC: Direct and Indirect Probes

Dave Lewis

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

I, Dave Lewis, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date:

Details of collaboration and publications:

ATLAS Collaboration, Study of the rare decays of B_s and B_0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector. *Eur. Phys. J. C.*, 76(513), 2016

ATLAS Collaboration, Search for a scalar partner of the Top Quark in the Jets+E_T^{miss} final state with the ATLAS detector. Moriond EW 2017, ATLAS-CONF-2017-020

This work was supported by the Science and Technology Facilities Council

Abstract

This thesis presents the results for two searches for new physics performed with the ATLAS experiment. The first, a search for the rare B-meson decay $B_s \to \mu\mu$ and measurement of its branching ratio, uses 25 fb^{-1} of $\sqrt{s} = 7$ and 8 TeV data recorded during 2011 and 2012. After observing a small number of these decays, a branching ratio of $\mathcal{B}(B_s \to \mu\mu) = (0.9^{+1.1}_{-0.8}) \times 10^{-9}$ is measured, assuming non-negative event yields. This is compatible with the Standard Model at the 2σ level. The second, a search for direct pair production of the supersymmetric top quark partner, is performed using 36.07 fb^{-1} of $\sqrt{s} = 13$ TeV data recorded during 2015 and 2016. Final states with a high jet multiplicity, no leptons and large missing transverse momentum are selected to target these decays, with several signal regions designed to cover a wide range of particle masses. No excess is observed, with all signal regions being compatible with the Standard Model within 2σ. Limits are set on the stop mass, excluding up to $m_{\tilde{t}_1} = 940$ GeV for values of $m_{\tilde{\chi}^0_1}$ below 160 GeV, assuming a 100% branching fraction to $\tilde{t}_1 \to t\tilde{\chi}^0_1$ decays. In addition two reinterpretations of this data are presented, for a gluino-mediated stop production scenario and a direct dark matter production scenario. No excess is observed for either model, and limits are set on the mass of the relevant particles. Finally a viability study into using machine learning techniques to improve on existing SUSY search methods has been performed, with the initial results proving promising.

Contents

Chapter 1

Introduction

From crystals to molecules, to atoms and beyond, science has always tried to answer two questions: what is the universe made of, and how does it interact. The last fifty years has brought incredible advances in both technology and scientific knowledge, both of which reach a pinnacle with the Large Hadron Collider (LHC). By colliding the building blocks of matter at an unprecedented energy, particle physicists are able to recreate the conditions of the universe mere billionths of a second after the Big Bang. By studying the remnants of these collisions, those two fundamental questions may be answered as scientists probe the very smallest scales of the universe.

Particle physicists use a robust theory called the Standard Model to understand how sub-atomic particles interact, and so far its predictions have been successful. That is not to say that everything is known, however - several glaring discrepancies between the Standard Model and both astrophysical and collider data exist, that cannot be explained by the Standard Model alone. Many theories have been proposed that could explain, or at least cover, those holes, but so far none have been corroborated by experimental data. The Large Hadron Collider may finally provide the evidence to prove that New Physics exists, and to bring us ever closer to understanding the universe.

This thesis contains the results of two searches for this evidence, performed by the ATLAS experiment on the LHC - one a direct search for decays of new particles, and the other an indirect probe, using a measurement of a Standard Model process to look for discrepancies. These two methods together may point to the nature of any New Physics, should it exist within reach of the LHC. The thesis is organised as follows: Chapter 2 details the Standard Model, the flaws with its assumptions, and how the theory Supersymmetry may solve its problems. Chapter 3 gives an overview of the ATLAS detector itself, explaining how it records and reconstructs the high-energy collisions, and Chapter 4 contains a measurement of the Lorentz Angle in the ATLAS Semiconductor Tracker, used to calibrate the detector. Chapter 5 is the first of the two New Physics probes, a measurement of the branching

ratio of the rare $B_s \to \mu\mu$ decay, a Standard Model process which could be sensitive to the existence of new, heavy particles. This measurement uses 25 fb^{-1} of data recorded by the ATLAS experiment during $\sqrt{s} = 7$ and 8 TeV collisions in 2011 and 2012. The second, direct probe of New Physics, a search for hadronic pair production of the supersymmetric top quark partner, is detailed in Chapter 6, using 36.07 fb^{-1} of $\sqrt{s} = 13$ TeV data recorded during data-taking in 2015 and 2016. Finally, Chapter 7 gives examples of reinterpretations that could be done to extend the use of available data, beyond the hadronic pair production search in Chapter 6, as well as a study into the viability of using machine learning to improve on the current search techniques. A summary and some concluding remarks are given in Chapter 8.

Chapter 2

The Standard Model and Beyond

Before any experiment can be carried out, the underlying theory must be considered. In the case of particle physics, this theory is called the Standard Model of particle physics (SM). The SM provides a description of the fundamental particles and the forces through which they interact, and was developed through both theoretical predictions and experimental observations. Robust as it is, standing up to 50 years of experimental scrutiny, there are several 'known unknowns' that motivate further experiments and theoretical development. This chapter provides an overview of the SM and its strengths, before exploring its flaws and one popular theory (amongst many) that aims to fix those flaws, called supersymmetry.

2.1 The Standard Model of Particle Physics

The Standard Model [1, 2] of particle physics summarises all known matter and anti-matter in the universe, and how they interact on a subatomic scale. Formulated as a Quantum Field Theory (QFT), particles are described as excitations of particle fields interacting through three fundamental forces - the weak, strong and electromagnetic forces. These forces are mediated by gauge bosons - for example, an electromagnetic interaction between two particles is equivalent to the particles exchanging a photon, the gauge boson of the electromagnetic force.

The Standard Model Particles

The SM divides particles into two categories, based upon a quantum number intrinsic to each particle called *spin (s)*, a quantum analogue to angular momentum. Fundamental particles whose spin is one-half are called elementary fermions and constitute the 'matter' in the universe. Fundamental particles

with integer spin values are called gauge bosons and are the mediators through which the elementary fermions interact with the fundamental forces. Fermions are then further divided into two types based on whether or not they interact with the strong force - those that do are called quarks, those that do not are called leptons.

Elementary Fermions

Any particle that is characterised by Fermi-Dirac statistics is called a fermion, which includes composite particles such as atoms and nuclei. These composite particles are not explicitly described by the SM, as they are formed with elementary fermions - in general, throughout this thesis, the term fermion will refer to only these elementary fermions.

The group of elementary fermions consists of six quarks and six leptons. The quarks are divided into two types, *up*-type and *down*-type quarks, where *up*-type quarks have positive $(+\frac{2}{3})$ electric charge and *down*-type quarks have negative $(-\frac{1}{3})$ electric charge. In addition there are also three generations of quarks, which differ only by their mass. As quarks interact through the strong force, they each carry a 'colour charge' quantum number - red, green or blue - which is conserved during interactions. Therefore each quark is actually a 'colour-triplet', or three indistinguishable states differing only by their colour charge. The leptons are similarly divided into two types based on their electric charge - those with charge (± 1), interacting through both the weak force and the electromagnetic force, and those without charge that interact only through the weak force. The neutral leptons (neutrinos) are assumed to be massless in the SM, though recent experimental observations have shown them to have a non-zero (albeit very small) mass. Finally, each fermion has an anti-particle equivalent that is identical except for a conjugation in the internal quantum numbers, such as the flavour and charge quantum numbers, colour for quarks and lepton number for leptons. The SM is ambiguous as to whether the neutrinos have separate anti-particle partners, or whether they are their own anti-particle - experimental evidence is required to make this distinction. Assuming they are not their own anti-particle, the two are identical except for chirality, defined as the projection of a particle's spin onto its direction of motion; neutrinos are left-handed (spin and momentum are in the same direction) while anti-neutrinos are right-handed (opposite directions).

Gauge Bosons

For each of the three forces in the SM - the weak and strong forces, and electromagnetism - there are gauge bosons that carry that force. Electromagnetism is arguably the most well-known of these forces and is carried by the photon. Electric charge is conserved during all particle interactions, and the photon has no mass or charge - it travels at the speed of light, c, and has infinite range. The

Generation	Name	Symbol	Mass [MeV]	Charge [e]
1st	Up	u	2.3	$+2/3$
	Down	d	4.8	$-1/3$
2nd	Charm	c	1.28×10^3	$+2/3$
	Strange	s	95	$-1/3$
3rd	Top	t	1.73×10^5	$+2/3$
	Bottom	b	4.18×10^3	$-1/3$

Table 2.1: Names and properties of the 6 SM quarks. Masses taken from [3].

Generation	Name	Symbol	Mass [MeV]	Charge [e]
1st	Electron	e	0.511	-1
	Electron neutrino	ν_e	$\leq 2 \times 10^{-3}$	0
2nd	Muon	μ	105.7	-1
	Muon neutrino	ν_μ	$\leq 2 \times 10^{-3}$	0
3rd	Tau	τ	1.78×10^3	-1
	Tau neutrino	ν_τ	$\leq 2 \times 10^{-3}$	0

Table 2.2: Names and properties of the 6 SM leptons. Masses taken from [3].

strong force is carried by gluons, particles with a colour charge but no mass or electromagnetic charge. Quarks are the only elementary fermions that interact with gluons, though unlike photons, gluons can also self-interact. Also unlike photons, the strong force has a limited range despite having a massless gauge boson, leading to an asymptotic freedom - as the interaction distance increases, the strength of the interaction increases, causing quarks to always be found in groups rather than individually. This is studied in more detail in Section 2.2. Finally, the weak force has three gauge bosons, the charged $W^\pm$ and the neutral Z bosons. All fermions can interact via the weak force, though charge is still conserved. Unlike the previous two forces, the weak force-carrying bosons have mass, meaning the weak force has a limited range, governed by the lifetime of its gauge bosons. All gauge bosons that mediate a force in the SM are vector bosons, with spin quantum number $J = 1$. In addition, the SM includes the Higgs boson, a massive neutral boson that doesn't carry a force, but instead interacts with particles to give them mass; the fact that the Higgs boson itself has mass means it also self-interacts. Unlike the other gauge bosons in the SM, the Higgs boson is a scalar boson, with spin $J = 0$ - this was an important quantity to determine when it was first discovered, as any other value of spin would not have been compatible with the SM [4].

	Boson	Symbol	Mass [GeV]	Charge [e]
Electromagnetism	Photon	γ	0	0
Strong force	Gluon	g	0	0
Weak force	$W^{\pm}$	-	80.39	± 1
	Z	-	91.19	0
Higgs field	Higgs	H	125.7	0

Table 2.3: Names and properties of the gauge bosons. Masses taken from [3].

2.2 Lagrangians and Gauge Groups

Mathematically, the particle content of the Standard Model and their interactions are described with a Lagrangian ($\mathcal{L}_{SM}$), which is invariant under gauge transformations. The local symmetry group, $SU(3) \times SU(2) \times U(1)$ describes the three forces present in the SM - $SU(3)$ describes the strong sector and Quantum Chromodynamics (QCD), $SU(2)$ describes the weak force, and $U(1)$ describes electromagnetism and Quantum Electrodynamics (QED). The two groups $SU(2)$ and $U(1)$ may be combined into the $SU(2) \times U(1)$ group, also known as the Electroweak sector (EWK), in a process called electroweak unification.

$\mathcal{L}_{SM}$ itself is formed of three terms:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EWK} + \mathcal{L}_{Higgs}. \tag{2.1}$$

Quantum Electrodynamics

The simplest QFT of the Standard Model, QED arises from the Abelian $U(1)$ group and describes electromagnetic interactions between charged particles and the photon. Its Lagrangian, $\mathcal{L}_{QED}$, is given by:

$$\mathcal{L}_{QED} = \bar{\psi}(i\not{D}_{QED} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.2}$$

where the ψ field spinor denotes the charged fermion, with mass m and charge Q, while F is the electric field tensor expressed in terms of the photon vector field A, $F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu}$. The covariant derivative $\not{D} = \gamma^{\mu}(\partial_{\mu} + iQA_{\mu})$ is used instead of the partial derivative ∂ to keep local gauge invariance. As the gauge group commutes, there is no interaction term between photons, meaning the only interactions are between a photon and a charged fermion.

Quantum Chromodynamics

QCD [5] comes from the $SU(3)$ symmetry group, and describes the interactions between colour particles (i.e. quarks and gluons), or 'partons' for short-hand. The QCD Lagrangian, $\mathcal{L}_{QCD}$, has a similar formalism to the QED Lagrangian, namely an interaction term between a fermion field and a boson gauge field plus a field tensor term, except in QCD the gluon field tensor $G^{\alpha}_{\mu\nu}$ replaces the electric field tensor $F^{\mu\nu}$, and is summed over all quark flavours (n = u,d,...):

$$\mathcal{L}_{QCD} = \sum_{n} \bar{\psi}^{b}_{n}(i\not{D}_{QCD} - m_n)\psi^{b}_{n} - \frac{1}{4}G^{\alpha}_{\mu\nu}G^{\alpha\mu\nu}. \tag{2.3}$$

The three colour charges (R, G, B and $\bar{R}$, $\bar{G}$, $\bar{B}$ for quarks and anti-quarks respectively) are represented by the index b, and the gluon index α runs from 1 to 8, one for each of the $(N^2_{\text{colours}} - 1)$ gluon vector fields, derived from the algebra of the $SU(3)$ group symmetry. The colour states of these gluons is given by the Gell-Mann matrices, and are independent of both each other and the colour singlet state. The covariant derivative $\not{D}_{QCD}$ and gluon tensor field are given below:

$$\not{D}_{QCD} = \gamma^{\mu}(\partial_{\mu} + ig_s T^{\alpha} A^{\alpha}_{\mu}) \tag{2.4}$$

$$G^{\alpha}_{\mu\nu} = \partial_{\mu}A^{\alpha}_{\nu} - \partial\nu A^{\alpha}_{\nu} - g_s f^{ijk} A^{j}_{\mu} A^{k}_{\nu}. \tag{2.5}$$

Similarly to QED, A^{α}_{μ} represents the gluon gauge field. Comparing the definition of $\not{D}_{QCD}$ to $\not{D}_{QED}$, its obvious that the QCD equivalent of a particle's charge, the colour charge, is given by $g_s T^{\alpha}$, where g_s is the strong coupling constant and T^{α} are the 8 generators of the $SU(3)$ group. The commutation relation of these generators also gives rise to the $SU(3)_C$ structure constants f^{ijk} (due to it being a non-Abelian group): $[T^i, T^j] = if^{ijk}T^k$. The extra term in $G^{\alpha}_{\mu\nu}$, with a dependence on g_s and f^{ijk}, allows for self-interaction between gluons.

This self-interaction with the gluon gauge field leads to one of the most important features of the strong interaction. Unlike in QED where the field strength has a $1/R^2$ relation with distance, the strong field is asymptotically free, increasing in strength as the interaction distance increases. At some point, then, it becomes energetically favourable to produce new quarks from the input energy, rather than using the energy to pull apart the parent quarks. This leads to a phenomenon known as quark confinement - it is not possible to observe a non-paired quark at energies below the QCD scale, Λ_{QCD}, as they hadronise to form bound states.

Two main groups of hadrons exist - baryons, a three quark bound state, and mesons, a quark-antiquark paired state - though recent results from the BES-III, Belle and LHCb experiments appear to have

discovered tetra- and even penta-quark bound states [6–9]. The effect of hadronisation can be seen in detectors as a roughly colinear shower of baryons and mesons, also known as a *jet* [5].

The Weak Force and Electroweak Unification

The weak force is described by the $SU(2)$ symmetry group [1], and is primarily known as the force responsible for nuclear β decays. Unlike the strong and electromagnetic forces, weak interactions allow the changing of quark flavour and violate parity by only acting on left-handed particles. $SU(2)$ is invariant in the weak isospin space (T, though the third component T_3 is the most important as it is conserved). In addition, $SU(2)$ is a non-Abelian group, meaning the gauge fields $W_\mu^{\alpha=1,2,3}$ may couple to one-another.

The Standard Model does not contain a Lagrangian for the weak force alone - instead, at an energy 100 GeV, the electromagnetic and weak forces unify. The unified electroweak field is described by the $SU(2) \times U(1)$ gauge group, with the weak hypercharge quantum number, Y, containing both the charge and weak isospin, $Q = T_3 + \frac{Y}{2}$. The electroweak Lagrangian ($\mathcal{L}_{EW}$) is then given by:

$$\mathcal{L}_{EW} = \sum_f \bar{\psi}_f i \not{D}_{EW} \psi_f - 1/4 W_\alpha^{\mu\nu} W_{\mu\nu}^\alpha - 1/4 B^{\mu\nu} B_{\mu\nu} \tag{2.6}$$

where ψ_f are the chiral fermion fields, summed over all elementary fermions, and the covariant derivative $\not{D}$ is given in terms of the $SU(2)$ and $U(1)$ gauge fields W^α and B, their gauge coupling constants g and g' respectively, and the $SU(2)$ generators τ_L^i:

$$\not{D}_{EW} = \gamma^\mu (\partial_\mu + i g' \frac{Y}{2} B_\mu + i g \frac{\tau_L^i}{2} W_\mu^i). \tag{2.7}$$

The addition in the Lagrangian of terms related to W^α and B describe the self interaction of the bosons. However the physical interpretation of the gauge bosons is a series of mixed states, with some mixing angle θ_W:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp i W_\mu^2), \tag{2.8}$$

$$Z_\mu = B_\mu^0 \cos(\theta_W) + W_\mu^3 \sin(\theta_W), \tag{2.9}$$

$$A_\mu = -B_\mu^0 \sin(\theta_W) + W_\mu^3 \cos(\theta_W), \tag{2.10}$$

where $W_\mu^\pm$ represents a charged massless state, and Z_μ and A_μ represent neutral massless states, with the $W^\pm$ bosons, Z boson and the photon as the corresponding gauge bosons. This mixing angle θ_W

is not given a value within the Standard Model, though its definition $\tan\theta_W = g'/g$ is given. Instead, θ_W is found experimentally, via the $\sin^2(\theta_W)$ quantity measurable in Z boson interactions and neutral-current processes [10], and is input into the SM. It is important to note at this point that none of these states include a mass term, nor is there a term in the Lagrangian by which these bosons may gain a mass, yet experiments show that the $W^{\pm}$ and Z bosons are in fact massive. For bosons to be massive, the electroweak symmetry has to be broken.

2.3 Symmetry Breaking and the Higgs Mechanism

If only the QCD and EW Lagrangians are taken into account, then $\mathcal{L}_{SM}$ contains no mass terms and all particles are massless. Therefore there must be an additional term in the SM that acts to give the fermions and weak bosons a mass.

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{mass}. \tag{2.11}$$

If the mass terms are added by hand, the gauge symmetry of the model is broken (the Dirac mass term $-m\bar{\psi}\psi$ is not invariant under electroweak symmetry). There must be some mechanism by which an initially massless particle may acquire a mass. As it turns out, the spontaneous breaking of the $SU(2) \times U(1)$ local gauge symmetry has the desired effect, as discovered by Higgs, Brout and Englert [11,12]. The Higgs mechanism introduces a new $SU(2)$ doublet, ϕ, of complex scalar fields, with weak hypercharge $Y_W = \pm 1/2$. This new doublet is invariant under the global $SU(2)$ transformation

$$\phi \rightarrow \phi' = e^{i\alpha_a \tau_a/2}\phi. \tag{2.12}$$

The Lagrangian of ϕ is obtained by adding a kinematic term, a mass term, and a self-interaction term with coupling strength λ:

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2. \tag{2.13}$$

A field potential, $V(\phi)$, can be defined as follows:

$$V(\phi) = \mu^2 |\phi|^2 + \lambda |\phi|^4. \tag{2.14}$$

If $\mu^2 < 0$ and $\lambda > 0$, this potential forms the classic 'wine bottle' potential, shown in Figure 2.1. In the Argand plane, the minimum of the potential forms a circle of radius $|\phi_0| = \sqrt{\frac{-\mu^2}{\lambda}} = \nu$. This

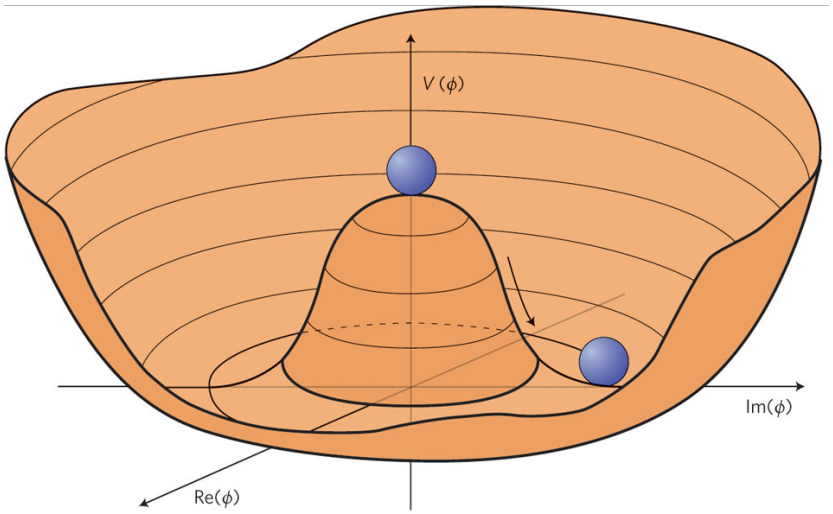

Figure 2.1: The Higgs potential, $V(\phi)$, shown in the Argand plane. The lowest energy point is randomly chosen from around the bottom of the 'bottle'. [2]

parameter ν is called the vacuum expectation value - if this value is non-zero, the local gauge symmetry is broken (which can be seen by eye, by looking at the potential - globally, it is symmetric, but from an arbitrary point on the minimum it no longer appears symmetric). Due to gauge invariance, the expansion $|\phi(x)| = \sqrt{\frac{1}{2}}(v + h(x))$, where $h(x)$ is the scalar Higgs field, can be substituted into the Lagrangian in Eq. 2.13:

$$\mathcal{L} = \left|\frac{1}{2}\left(\partial_\mu(v+h) + ig\frac{1}{2}\tau \cdot W_\mu(v+h)\right)\right|^2 - V(\phi) - \frac{1}{4}W_{\mu\nu} \cdot W^{\mu\nu}. \tag{2.15}$$

Parameterising fluctuations from the vacuum ϕ_0 in terms of the Pauli matrices leads to the appearance of four real fields, θ_1, θ_2, θ_3, and the Higgs field h:

$$\phi = \sqrt{\frac{1}{2}}\begin{pmatrix} \theta_2 + i\theta_1 \\ \nu + h - i\theta_3 \end{pmatrix}. \tag{2.16}$$

The three fields $\theta_{1,2,3}$ are independent, and fully parameterise the deviations from the vacuum ϕ_0. Now, the Lagrangian is locally $SU(2)$ invariant, and the three massless Goldstone boson fields $\theta_{1,2,3}$ can be gauged, removing any trace of $\theta(x)$. Substituting $\theta_{1,2,3}$ back into the Lagrangian leads to mass states for physical gauge bosons in terms of the gauge fields W_μ^α and B_μ [1], leading to three massive gauge fields and one massless gauge field, below, and one massive scalar, h:

$$W_\mu^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}}, \text{ with } m_W = \frac{1}{2}\nu g, \tag{2.17}$$

$$Z_\mu^0 = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}}, \text{ with } m_Z = \frac{1}{2}\nu\sqrt{g^2+g'^2}, \tag{2.18}$$

$$A_\mu^0 = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2+g'^2}}, \text{ with } m_A = 0, \tag{2.19}$$

$$\cos\theta_W = \frac{m_W}{m_Z}, \tag{2.20}$$

$$m_h = \sqrt{2\lambda\, v^2} = \sqrt{2}\mu, \tag{2.21}$$

where A is the photon field as per QED - hence why the photon is massless. The mass term m_h, plus the kinetic term and the cubic and quartic terms present when equation 2.15 is fully expanded, represent a new particle, the Higgs boson. The 3rd and 4th order terms in Eq. 2.15 describe 3- and 4-point interactions in h, meaning self-coupling is possible, explaining why the mass-giving particle itself has mass. This new particle was discovered in 2012 by the ATLAS and CMS experiments, with a mass $m_h = 125$ GeV [13, 14].

The fermions also gain mass through this mechanism, via interactions between the Higgs field and the scalar fermion fields, called Yukawa couplings. This is only made possible by spontaneous symmetry breaking, as a fermion mass term $-m\bar{\psi}\psi$ is excluded in Eq. 2.6 by gauge invariance. For example, to generate the electron mass, substitute $\phi = \sqrt{\frac{1}{2}}(v + h(x))$ into the Lagrangian, thereby spontaneously breaking the symmetry and leading to a Lagrangian:

$$\mathcal{L}_3 = -\frac{G_e}{\sqrt{2}}v(\bar{e}_L e_R + \bar{e}_R e_L) - \frac{G_e}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L)h, \tag{2.22}$$

where G_e is the Yukawa coupling between the electron and the Higgs field. This coupling can be chosen so that:

$$m_e = \frac{G_e v}{\sqrt{2}} \tag{2.23}$$

which generates the required electron mass:

$$\mathcal{L}_e = -m_e\bar{e}e - \frac{m_e}{v}\bar{e}eh. \tag{2.24}$$

A similar process can be followed for the quarks, too, with couplings $G_{u,d}$. However, these Yukawa couplings are arbitrary, and as a result, the fermion masses are not predicted by the SM. Empirical

input is required to resolve this - in particular, the neutrino is problematic as experimental results indicate it has a non-zero mass [15–17], while the Higgs mechanism states this is only possible if a right-handed neutrino exists, for which no evidence has been found.

The final Lagrangian for the Standard Model, then, looks like this:

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_{Higgs}, \tag{2.25}$$

with $\mathcal{L}_{Higgs}$ containing both the electroweak boson terms and the fermionic Yukawa coupling terms.

2.4 Beyond the Standard Model

The SM has held up to rigorous testing over the years, as well as predicting the existence of particles decades before they were discovered. So far there is no direct evidence that the model is wrong - however, it does not tell the full story. The arbitrary fermion Yukawa couplings have been previously discussed, and the implications for neutrinos, but that isn't the only issue with the SM. Gravity, a force that so obviously exists at large scales, is not included in the Standard Model and, if compared to the other three forces, is $O(10^{37})$ times weaker. Certainly, the SM will need to be extended to cover physics at the Planck Scale ($M_P = 2 \times 10^{18}$ GeV) where gravity is expected to unify with the other forces. In addition, astrophysical and cosmological observations indicate that only 5% of our universe is comprised of the baryonic matter of the SM, with 25% made of an unknown type of matter called Dark Matter and the remaining 70% being an even more nebulous quantity known as Dark Energy, neither of which the SM offers any explanation to.

A large problem with the SM is the 'Hierarchy Problem', which relates to the Electroweak scale and loop corrections to Higgs interactions. In theory the SM should be valid at infinite energy scales, but in practice it breaks down at the Planck Scale, or at least at a New Physics scale Λ where additional particles and interactions enter the equation. At leading order, this isn't so much of an issue as the SM is a fully renormalisable theory, meaning that as $\Lambda \to \infty$, μ^2 can be chosen such that the scalar mass term has the desired value. When loop diagrams are considered, as in Figure 2.2, the mass of the Higgs has to be corrected by Δm_H^2, leading to a relationship that depends on Λ and the fermionic coupling strength λ_f: [18, 19]

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda^2 + \dots. \tag{2.26}$$

As Λ increases up to the Planck Scale, the corrections arising from additional loop diagrams increase until Δm_H^2 is ~ 30 orders of magnitude larger than the Higgs mass itself. This Hierarchy Problem

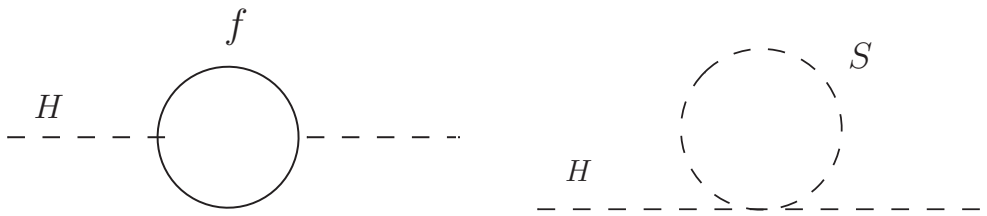

Figure 2.2: *Left:* Correction to the Higgs mass due to a fermion loop. *Right:* Correction to the Higgs mass due to a new-physics scalar field S.

only applies to the Higgs, as the fermion and boson masses do not have the same sensitivity to Λ as m_H; however the entire Standard Model spectrum is indirectly sensitive to these corrections as the particle masses are related to the Higgs mass, and hence are also sensitive to Λ.

Assuming that there are some new heavy particles at the Λ scale, well above the electroweak scale, there are two options for dealing with these quadratic divergences: one, the new particles do not couple even remotely to the Higgs and gain their mass through some unknown process, or two, there is a cancellation between the various contributions to the corrections. Simplicity dictates the second option is the more likely - all that is required is an additional symmetry.

If each fermion in the SM is accompanied by two complex, massive, scalar particles S, that couple to the Higgs with strength $\lambda_S = |\lambda_f|^2$, then the Λ^2 contributions to Δm_H^2 arising from fermion and boson loops can be exactly cancelled. This implies a fundamental relationship between the SM fermions and these new bosons, also called a *supersymmetry*, or SUSY. An operator, Q, can transform one into the other:

$$Q|Boson\rangle = |Fermion\rangle, \quad Q|Fermion\rangle = |Boson\rangle. \tag{2.27}$$

Under this kind of transformation, the *sparticles* (S) would have the same quantum numbers (except spin, differing by $\pm 1/2$) and mass as their partner particles. However no such particles have been found, so there must be a broken symmetry pushing the sparticle masses upwards. This symmetry breaking must be 'soft', to avoid the sparticle masses becoming so heavy that the loop contributions no longer cancel. The actual mechanism is unknown, though several models have been studied - this thesis assumes that the symmetry breaking is *gauge-mediated*, whereby electroweak and strong interaction loops relate a *hidden sector* of fields to the SM fields through *messenger* particles [18]. By assuming $\Lambda \sim M_P$ and $\lambda_S \sim 1$, the SUSY mass scale can be estimated - the lightest sparticles should have mass of order 1 TeV. In other words, those light SUSY particles could be discovered at the LHC.

Minimal Supersymmetry

The Minimal Supersymmetric Standard Model (MSSM) [20] is a popular version of SUSY that contains the minimum number of additional particles required to solve the Hierarchy Problem. For each SM fermion, there are two spin-0 sparticles (notationally left- and right-handed, based on the helicity of the partner fermion) called sfermions, while the SM bosons have fermionic sparticle partners called gauginos. The Higgs gains two additional doublets, leading to 5 total Higgs particles including the 125 GeV SM Higgs. Sparticles are denoted with a tilde above the particle symbol, for example $\tilde{t}$, $\tilde{\mu}$, or $\tilde{\nu}$, with the naming scheme as follows: partners of SM fermions keep the same name with an 's' in front (for example, the top quark has a partner called the stop), while SM bosons have partners with names ending in 'ino'. The photino, gluino, wino, bino and higgsino are the partners of the SM bosons, with the electroweak partners mixing to form the observable mass states, the neutralino and chargino, as shown in Table 2.4. The MSSM also predicts the existence of the gravitino, which implies the existence of the graviton, a particle that would mediate gravitational interactions and provide an explanation for why gravity appears to be a much weaker force than the other fundamental forces.

Table 2.4 shows the particle content of the MSSM. For some sparticles, mixing occurs between the gauge eigenstates, producing different mass eigenstates which are observable. For sfermions, the mixing is proportional to the mass of the SM particles, meaning the mixing is only non-negligible for the heaviest 3^{rd} generation particles. The mixing of the gauginos is governed by the matrices:

$$\begin{pmatrix} \tilde{\chi}_1^{\pm} \\ \tilde{\chi}_2^{\pm} \end{pmatrix} = \begin{pmatrix} M_2 & \sqrt{2} m_W \sin\beta \\ \sqrt{2} m_W \cos\beta & \mu \end{pmatrix} \begin{pmatrix} \tilde{W}^{\pm} \\ \tilde{H}^{\pm} \end{pmatrix} \tag{2.28}$$

$$\begin{pmatrix} \tilde{\chi}_1^0 \\ \tilde{\chi}_2^0 \\ \tilde{\chi}_3^0 \\ \tilde{\chi}_4^0 \end{pmatrix} = \begin{pmatrix} M_1 & 0 & -m_Z \cos\beta \sin\theta_W & m_Z \sin\beta \cos\theta_W \\ 0 & M2 & m_Z \cos\beta \cos\theta_W & m_Z \sin\beta \cos\theta_W \\ -m_Z \cos\beta \sin\theta_W & m_Z \cos\beta \cos\theta_W & 0 & -\mu \\ m_Z \sin\beta \cos\theta_W & m_Z \sin\beta \cos\theta_W & -\mu & 0 \end{pmatrix} \begin{pmatrix} \tilde{B}^0 \\ \tilde{W}^0 \\ \tilde{H}_u^0 \\ \tilde{H}_d^0 \end{pmatrix} \tag{2.29}$$

where M_1, M_2 are the masses of the bino and wino respectively, and μ is the higgsino mass. $\tan(\theta_W)$ is the ratio between electroweak coupling constants g and g', as per the SM, and $\tan(\beta)$ is the ratio of the vacuum expectation values of the two Higgs doublets, v_u and v_d, themselves constrained by the SM Higgs VEV, $v = \sqrt{v_u^2 + v_d^2} = \sqrt{2m_Z^2/(g^2 + g'^2)}$.

The MSSM, along with most other SUSY models, follows the concept of *naturalness* [21]. The idea is that because of the number of free parameters in the MSSM, there are a nearly infinite number of possible values for those parameters that would lead to the correct Higgs mass. A 'natural' SUSY

Name	Spin	Gauge eigenstates	Mass eigenstates
Squarks ($\tilde{q}$)	0	$\begin{pmatrix}\tilde{u}_L,\tilde{u}_R\\ \tilde{c}_L,\tilde{c}_R\\ \tilde{t}_L,\tilde{t}_R\end{pmatrix}, \begin{pmatrix}\tilde{d}_L,\tilde{d}_R\\ \tilde{s}_L,\tilde{s}_R\\ \tilde{b}_L,\tilde{b}_R\end{pmatrix}$	$\begin{pmatrix}-\\ -\\ \tilde{t}_1,\tilde{t}_2\end{pmatrix}, \begin{pmatrix}-\\ -\\ \tilde{b}_1,\tilde{b}_2\end{pmatrix}$
Sleptons ($\tilde{l}$)	0	$\begin{pmatrix}\tilde{e}_L,\tilde{e}_R\\ \tilde{\mu}_L,\tilde{\mu}_R\\ \tilde{\tau}_L,\tilde{\tau}_R\end{pmatrix}, \begin{pmatrix}\tilde{\nu}_e\\ \tilde{\nu}_\mu\\ \tilde{\nu}_\tau\end{pmatrix}$	$\begin{pmatrix}-\\ -\\ \tilde{\tau}_1,\tilde{\tau}_2\end{pmatrix}, \begin{pmatrix}-\\ -\\ -\end{pmatrix}$
Higgs bosons	0	$(H_u^0, H_d^0, H_u^+, H_d^-)$	$(h^0, H^0, A^0, H^\pm)$
Gauginos	1/2	$\begin{pmatrix}\tilde{B}^0\\ \tilde{W}^0\\ \tilde{H}_u^0\\ \tilde{H}_d^0\end{pmatrix}, \begin{pmatrix}\tilde{W}^\pm\\ \tilde{H}_u^+\\ \tilde{H}_d^-\end{pmatrix}$	$\begin{pmatrix}\tilde{\chi}_1^0\\ \tilde{\chi}_2^0\\ \tilde{\chi}_3^0\\ \tilde{\chi}_4^0\end{pmatrix}, \begin{pmatrix}\tilde{\chi}_1^\pm\\ \tilde{\chi}_2^\pm\end{pmatrix}$
Gluino	1/2	$\tilde{g}$	-
Gravitino	3/2	$\tilde{G}$	-

Table 2.4: Spin, gauge eigenstates and mass eigenstates of the supersymmetric partner particles. When not stated, the mass eigenstates are the same as the gauge eigenstates; otherwise, the mass eigenstates are mixed states of the gauge eigenstates.

model would thus aim to keep the amount of fine-tuning of the free parameters to a minimum. The definition of 'minimum' varies from model to model, but in general it requires the Lightest SUSY Particle (LSP, usually the neutralino χ_1^0) to be lighter than ~ 350 GeV, the 3^{rd} generation sfermions to be lighter than 1 TeV, and the gluino to have < 2 TeV mass [18, 21]. Heavier versions of these particles are possible, but it requires tuning the parameters of the model, stretching the idea of a natural theory. The requirements on the other sfermion generations and gauginos are less stringent, however if these other sparticles are much heavier an additional symmetry breaking mechanism is needed to explain this mass difference.

R-Parity

A consequence of the MSSM Lagrangian is that terms arise that allow violation of both baryon and lepton symmetries. This is intended, to a certain degree, to explain the fact that the universe is dominated by matter over antimatter, but these symmetry violations may be larger than is required

for purely this reason. If SUSY did cause baryon and lepton number violation to that degree, the lifetime of the proton would be much shorter than current measurements ($> 5.8 \times 10^{29}$ years) - therefore there must be protection against this. To remove the offending terms, a new symmetry is introduced, called 'R-Parity', defined as the quantum number:

$$R_P = (-1)^{3(B-L)+2s} \tag{2.30}$$

where B and L are the baryon and lepton numbers and s is the spin of the particle. SM particles have a value of $R_P = +1$, and SUSY particles have $R_P = -1$, meaning only interactions involving sparticles may cause violation of R_P, and hence the number of sparticles at each interaction vertex must be conserved. While there are models that contain R-Parity violating decays, they require an extra mechanism to prevent proton decay. This thesis only considers R-Parity conserving SUSY models.

An implication of conserving R-Parity is that the LSP cannot decay - decays to other SUSY particles are forbidden by conservation of energy, and decays to SM particles are forbidden by R-Parity. The LSP is thus a built-in Dark Matter candidate [22] - a heavy, weakly-interacting particle with a long lifetime, that would still be around from the beginning of the universe. This implication is one of the reasons why SUSY is such a popular model, as it does not require any additions to explain dark matter.

Supersymmetry at the LHC

As previously discussed, the concept of naturalness places constraints on the mass of several sparticles, requiring them to be relatively light. As a result they may be found for the first time at the LHC, as the centre-of-mass energy $\sqrt{s} = 13$ TeV places any sub-TeV particle within reach. The high luminosity expectation for the LHC also means that higher-mass particles (for example, the gluino) or interactions with very low cross-sections may also be observed. Figure 2.3 shows the predicted production cross-section vs particle mass for various SUSY particles. Two interesting points can be noted here; firstly, the electroweak SUSY particles, like the chargino, have extremely low cross sections at the LHC despite their low masses - the LHC is a proton-proton collider, so QCD interactions are dominant, while the gauginos interact mostly with leptons through electroweak processes. Secondly, while the gluino has the largest predicted cross-section, it is also expected to have the highest mass of the shown particles. In effect, this balances out, meaning a stop with sub-TeV mass has a similar expected cross-section at the LHC as a gluino with $\sim$ 1.5 - 2 TeV mass.

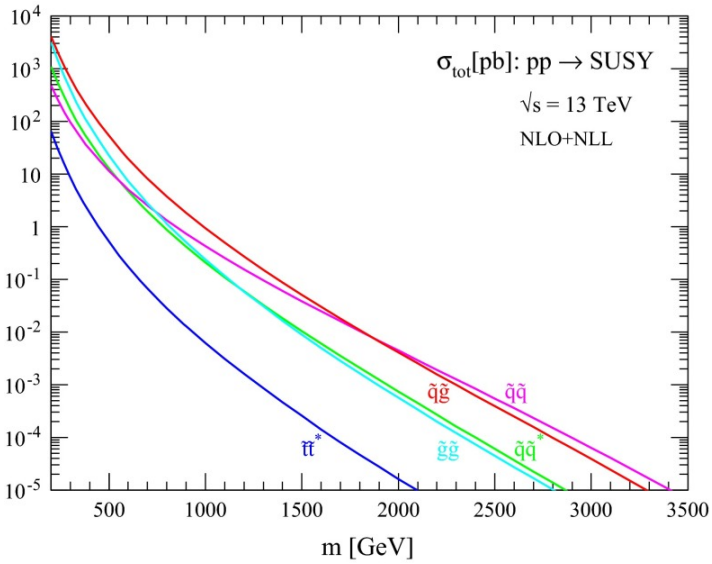

Figure 2.3: Predicted production cross sections as a function of sparticle mass, at $\sqrt{s} = 13$ TeV. Taken from [23].

Simplified Models

The MSSM has a huge number of free parameters - around 100, depending on the exact implementation. Other SUSY models such as the Constrained MSSM (cMSSM) or Phenomenological MSSM (pMSSM) reduce the number of free parameters by making assumptions and including experimental results, but there can still be many unknowns. Too many, in fact, to produce enough simulated events to make accurate predictions about the whole phase space in the model. Instead, analyses at the LHC use so-called simplified models, with a number of assumptions made to reduce the number of parameters to a manageable amount. Usually the only free parameters are the masses of the particles involved in the decay chain - two or three, at most. This leads to a much reduced phase space that can be explored with the LHC data, though any limits or observations have the caveat that the assumptions made may not reflect the reality of the situation. [24] Much theoretical effort goes into both creating new simplified models to experimentally search for different types of SUSY decays [25], and to reinterpreting existing experimental data that originally used a simplified model in terms of more complex theories [26–28] to constrain the model parameters.

Some examples of the Feynman diagrams for these simplified models are shown in Figure 2.4. All of these decay chains have a fixed number of interaction vertices, and at each the number of sparticles is conserved. The LSP - which throughout this thesis will be assumed to be the neutralino, χ_1^0 - does not decay, conserving R-Parity and appearing to the detector as missing energy. These decay chains

can then be simulated and looked for in data, exploiting the well-defined kinematics in the transverse plane of the system.

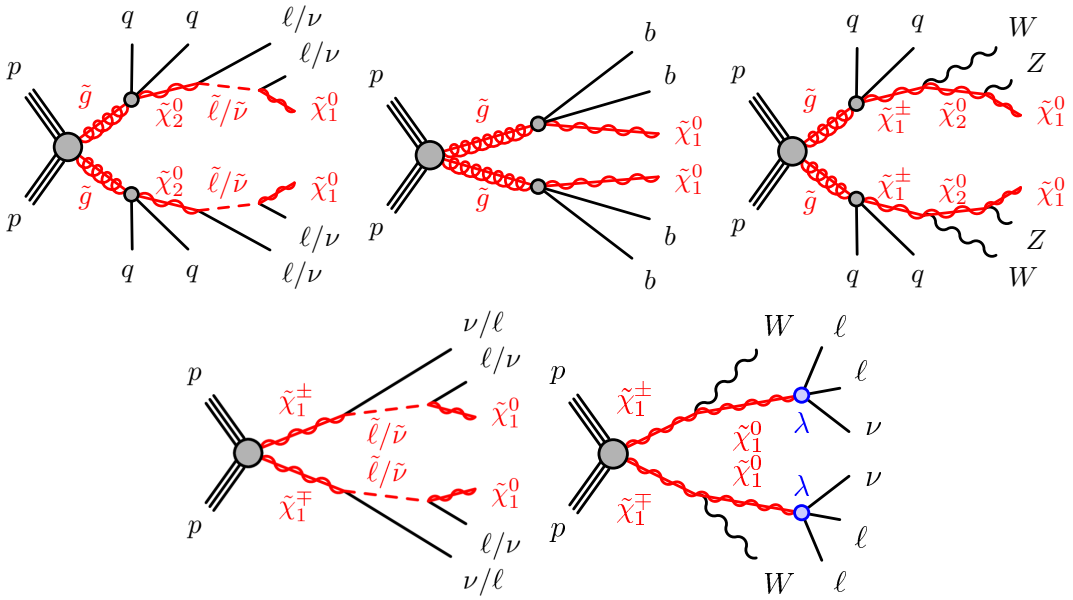

Figure 2.4: Examples of simplified SUSY models - note the well-defined decay chains that provide exploitable kinematics in the transverse plane. In most cases, the final decay products include a non-decaying LSP ($\tilde{\chi}^0$) that would appear as missing energy to a detector - the exception being the last model, where R-parity violation is allowed.

In general, the method of searching for a SUSY decay involves looking for events with large missing transverse energy, E_T^{miss}. This E_T^{miss} is provided by the $\tilde{\chi}_1^0$, which ends the decay chain of any sparticles, in order to conserve R-parity. Unless the $\tilde{\chi}_1^0$ is very heavy or highly boosted in the transverse plane, the missing energy alone is not enough to distinguish a SUSY decay from SM processes - which is where the decay chain of the simplified model is useful. By looking for E_T^{miss} in conjunction with the SM particles and the kinematics predicted by the simplified models, a SUSY signal may be observed. The invariant mass of the SM particles produced also helps calculate the mass of the parent sparticles, which may be difficult to find with the E_T^{miss} alone. Of course, using the E_T^{miss} as the key signature means that SM processes that involve neutrinos ($W \to l\nu$ or $Z \to \nu\nu$, for example) will start to look a lot like the SUSY signal - however, this is best left as a discussion for Chapter 6.

2.5 Rare B Decays

The problem with the simplified models used to search directly for new particles is that they make a lot of assumptions. If those assumptions turn out to be incorrect, then the new physics may never be found. Ideally there would be a method by which one can look indirectly for SUSY, without having to make a large number of assumptions that could lead to misleading conclusions. Thankfully the Standard Model provides for this - similarly to the loop corrections to the Higgs mass, shown in Figure 2.2, new heavy, virtual particles can be added into any loop process. If such a contribution exists, then the branching ratio for the process will deviate away from the SM prediction, as the new diagram contributes to the matrix element. While this sounds ideal, the problem is that if you choose to look at a common decay, the uncertainties on the tree-level process alone are larger than any effect the new physics may cause, and the new physics is essentially invisible. A rare process, that can only occur through loop diagrams with a much lower branching ratio and smaller uncertainties, would make the effect from new physics more visible. Choosing a much rarer process carries its own problems, though - it has to be seen before it can be measured.

Flavour Changing Neutral Currents

In the SM, quarks can change flavour through charged current interactions (FCCC) - top quarks, for example, decay to a bottom quark and a W boson. Changing flavour to a quark of the same type, with the same charge, is forbidden at tree-level in the SM and can only occur through loop diagrams. These Flavour Changing Neutral Currents (FCNC) are suppressed by the Glashow-Iliopoulos-Maiani (GIM) mechanism [29, 30]; the fact that quarks form doublets $\binom{u}{d}$, $\binom{c}{s}$, means that FCNC terms arising from Z couplings to, for example, $d\bar{s}$ are cancelled by the same terms from Z couplings to $s\bar{d}$. The GIM mechanism also predicted the existence of the charm quark, as it was required for the quarks to form doublets in the first place [29, 30].

FCNC interactions can still occur in the SM, however. By introducing a W loop interaction, the overall neutrality of the interaction can be preserved, while each interaction vertex is charged. Examples of these loop diagrams - called *penguin* or *box* diagrams depending on the number of virtual quarks - are shown in Figure 2.6. A classic example of a FCNC is the $B_d \to K^* \mu\mu$ decay - a B_d meson, consisting of a $\bar{b}d$ quark pair, decays to a kaon, $\bar{s}d$, via a W loop that emits a Z boson which decays to two muons (see Figure 2.5). In the Standard Model this has a branching ratio of $1.07 \pm 0.22 \times 10^{-6}$ [3], suppressed by both the GIM mechanism and CKM suppression, which defines the probability for a quark to change flavour. This decay is also an important one for indirectly probing new physics -the LHCb experiment sees an anomaly in one of the kinematic variables [31], P_5', a potential indicator of BSM physics [32], though the result from the CMS experiment is in agreement with the SM [33].

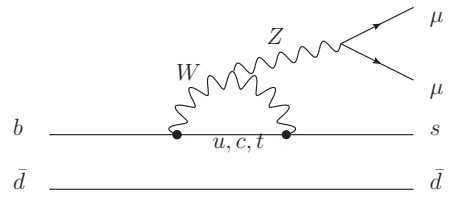

Figure 2.5: Penguin diagram for the $B_d \to K^* \mu\mu$ decay, an important example of a FCNC.

Rare Dilepton Decays as a New Physics Probe

Another example of a FCNC process is $B^0_{(s)} \to l^+ l^-$, where a $B^0_{(s)}$ meson decays to two charged leptons via a W boson loop diagram (see Figure 2.6). This decay undergoes both GIM suppression and CKM suppression, as the probability for a b-quark to decay directly to a lighter down-type quark is very small; but it also undergoes helicity suppression, by a factor $m_l^2/m^2_{B_{(s)}}$, driving the branching fraction even further down. In total, there are six decays of this kind - two mesons, with either a d or an s quark, that can both decay to all three lepton generations. A B_s decay to a given lepton will have a larger branching fraction than the same B_d decay; the muon channel has a larger branching fraction than the electron channel, and the τ channel branching fraction is larger than both. This would make the τ channel ideal to look for these rare processes in, however in practice, τs are hard to reconstruct, as they decay to a ν and a W, which can itself decay leptonically or hadronically. Instead, the muon channel is the best channel to look in, despite an order of magnitude smaller branching fraction, purely due to a final state that is easier to reconstruct. CMS and LHCb jointly published the first observation of $B_s \to \mu\mu$ decays using $\sqrt{s} = 8$ TeV LHC data, at a significance of 6.2σ [34], with LHCb following on with the first observation of $B^0_s \to \mu^+\mu^-$ by a single experiment using $\sqrt{s} = 13$ TeV LHC data [35]. The ATLAS result, using 25 fb^{-1} of $\sqrt{s} = 8$ TeV data, is covered in Chapter 5 of this thesis.

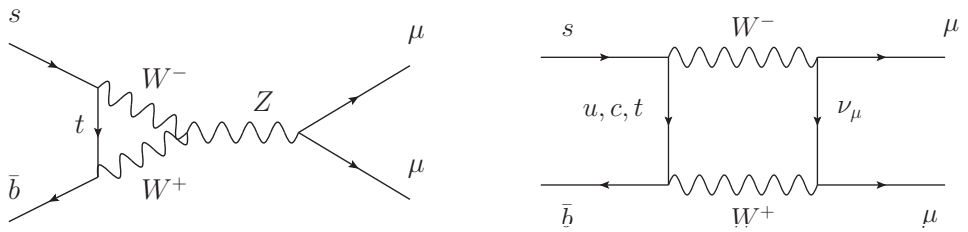

Figure 2.6: *Left:* A FCNC penguin diagram, as allowed in the Standard Model. *Right:* The box diagram version of the same process, $B_s \to \mu\mu$.

The advantage of the rarity of this decay is that if SUSY exists, additional loop diagrams involving sparticles (as in Figure 2.7) will cause the branching ratio to deviate away from the SM prediction.

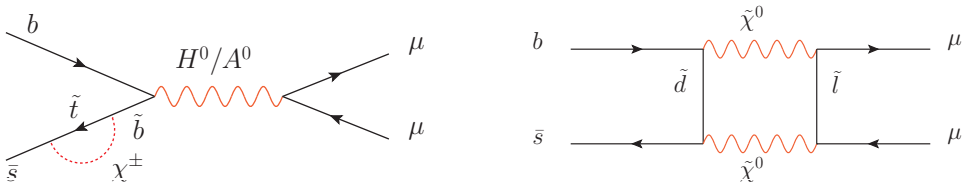

Figure 2.7: Feynman diagrams of possible new interactions caused by the addition of SUSY particles. These diagrams interfere with the Standard Model diagrams in Figure 2.6, and may cause the branching ratio of the $B_s \to \mu\mu$ decays to significantly deviate away from the SM prediction.

The theoretical precision of the decay means that this possible deviation could be seen with relatively few events - again, useful as the decay is exceedingly rare. As a probe for new physics, $B_s \to \mu\mu$ is very powerful; nearly all BSM models have some new particle or coupling that enhances the branching fraction for the decay, from heavy Z' resonances to new Higgs partners, sparticles and leptoquarks. In addition, with a direct search, the range of masses for new particles is limited to at most the centre of mass energy of the collider; with an indirect search, utilising loop diagrams, the mediator particle is virtual and thus its mass may far exceed what is possible to find with a direct search.

Focusing on SUSY, Figure 2.8 shows the branching ratio as a function of gaugino mass in the cMSSM, demonstrating the sensitivity to the SUSY scale. (The red line indicates the current measurement of the branching ratio from the CMS/LHCb combination, clearly excluding much of the cMSSM phase space). The enhancement in the branching ratio is of the order $\tan^3(\beta)$, where $\tan\beta$ is the ratio of vacuum expectation values of the two Higgs fields present in a SUSY model [36]. Similarly, the ratio between the $B_s \to \mu\mu$ and $B_d \to \mu\mu$ branching ratios is also sensitive to BSM physics, and can be used to exclude much of the SUSY phase space, as demonstrated by Figure 2.9.

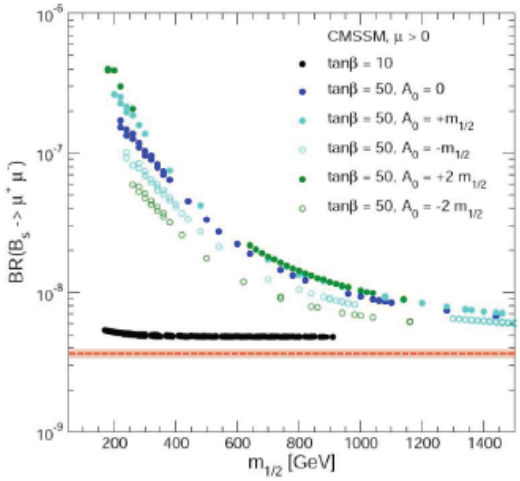

Figure 2.8: Predicted branching ratio in the CMSSM, with selected values of $\tan\beta$ and A_0, as a function of gaugino mass. Taken from [37]. The dashed red line indicates the branching ratio measured by CMS and LHCb [34].

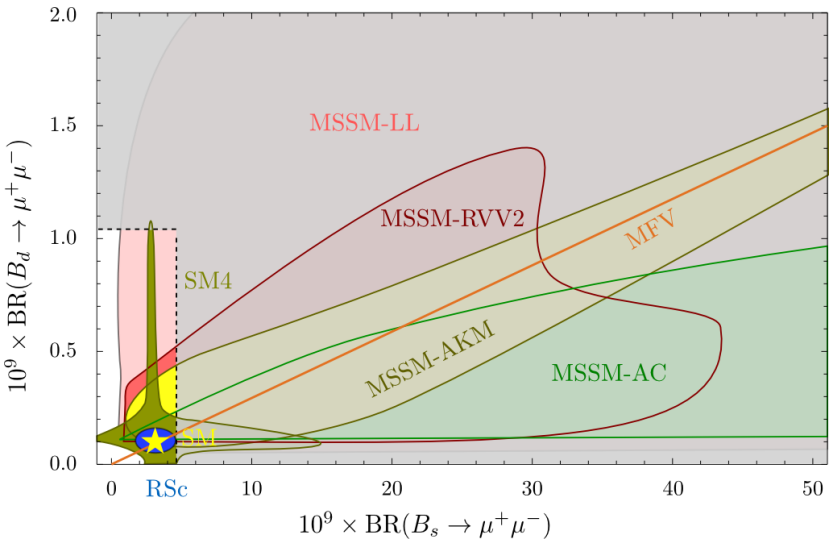

Figure 2.9: Correlation between the branching ratios for B_s and $B_d \to \mu\mu$, with the shaded areas the predictions made by various SUSY models. [38]. The grey area shows the excluded branching ratios, prior to the CMS and LHCb measurement.

Chapter 3

The LHC and the ATLAS Detector

3.1 The Large Hadron Collider

The LHC is a proton-proton synchrotron collider at CERN, near Geneva [39]. Its 27 km circumference tunnel crosses the French-Swiss border, between 50 and 175 m underground. The purpose is to accelerate bunches of protons up to energies of 7 TeV, then collide them together at centre-of-mass energies up to $\sqrt{s} = 14$ TeV. This is achieved by a complex of smaller accelerators, each feeding into larger, higher energy machines - a diagram of the whole CERN accelerator complex is shown in Figure 3.1. First, protons are fed into Linac2 and accelerated to 50 MeV, before being injected into the Proton Synchrotron Booster to reach 1.4 GeV of energy. After the Proton Synchrotron Booster the proton bunches are fed into the Proton Synchrotron and accelerated to 25 GeV, then into the Super Proton Synchrotron which propels them to 450 GeV. Finally the bunches are injected into the LHC as two beams going in opposite directions, and circulated until they reach the maximum beam energy of 7 TeV, at which point collisions may begin at the four interaction points.

Not only is the LHC designed to reach the highest centre-of-mass energy of any accelerator yet ($\sqrt{s} = 14$ TeV, compared to Fermilab's Tevatron at $\sqrt{s} = 1.96$ TeV [41]), but in order to discover extremely rare processes, it also has to deliver a large amount of data. The design instantaneous luminosity of the LHC is 10^{34} cm^{-2}s^{-1}, which is achievable by having bunches of 10^{11} protons collide every 25ns, corresponding to roughly 40 million p-p collisions per second. In 2012, while running at $\sqrt{s} = 8$ TeV, the LHC achieved a peak instantaneous luminosity of 7.7×10^{33} cm^{-2}s^{-1}, and an integrated luminosity of approximately 21 fb^{-1}. In 2015, running at $\sqrt{s} = 13$ TeV, the LHC reached a peak instantaneous luminosity of 5.0×10^{33} cm^{-2}s^{-1}, and delivered 4.2 fb^{-1} of integrated luminosity (shown in Figure 3.2). The design peak instantaneous luminosity was reached during the 2016 run, exceeding 10^{34} cm^{-2}s^{-1}, delivering nearly 40 fb^{-1} of pp collisions by the end of the year - this was possible by reducing the

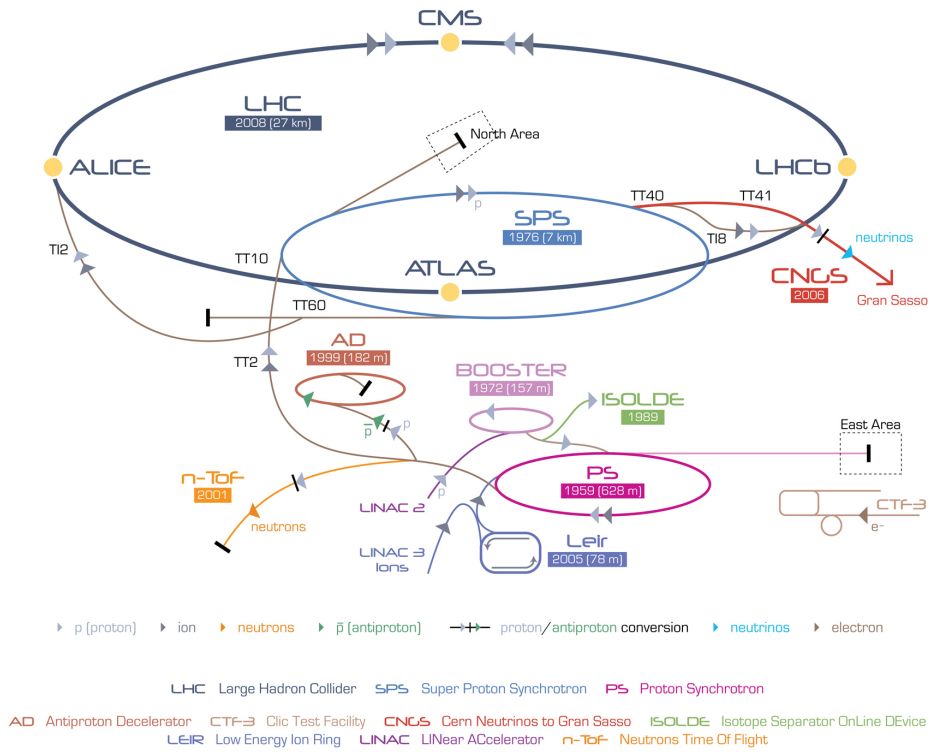

Figure 3.1: A representation of the CERN accelerator complex, starting with the LINAC2 and ending with the LHC. [40]

bunch spacing, the gap between proton bunches, to the design specification of 25 ns rather than the 50 ns spacing used previously.

3.2 The ATLAS Detector

One of the largest experiments at the LHC is A Toroidal LHC ApparatuS (ATLAS) [43]. ATLAS is a general-purpose experiment designed to study as many different physical processes as possible, from measuring common SM physics to discovering rare decays and searching for physics beyond the SM. A cylindrical detector, it provides a full 4π solid angle coverage for complete collision reconstruction. The detector itself is comprised of many layers; starting from the innermost layer, these are the pixel detector, the semiconductor tracker, the transition radiation tracker, the electromagnetic calorimeter, the hadronic calorimeter and the muon spectrometer. Each layer is described further in the corresponding subsections below. The inner detector, a combined term for the pixel detector, semiconductor tracker and transition radiation tracker, is inside a solenoid providing a 2 T magnetic field, while the muon

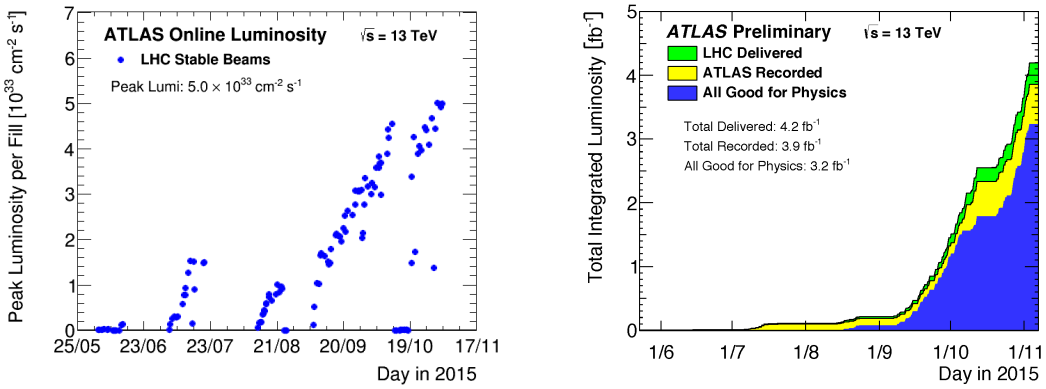

Figure 3.2: *Left:* The instantaneous luminosity of the LHC as a function of day during 2015 running. *Right:* The integrated luminosity delivered by the LHC and collected by ATLAS as a function of day, during 2015 running. [42]

spectrometer uses the detector's namesake, a toroidal magnet made of 8 spokes with a magnetic field strength varying between 0.5 and 1 T. Figure 3.3 gives a cut-away view of the detector, where each part of the detector is highlighted in a different colour, and an average-sized person is included for comparison.

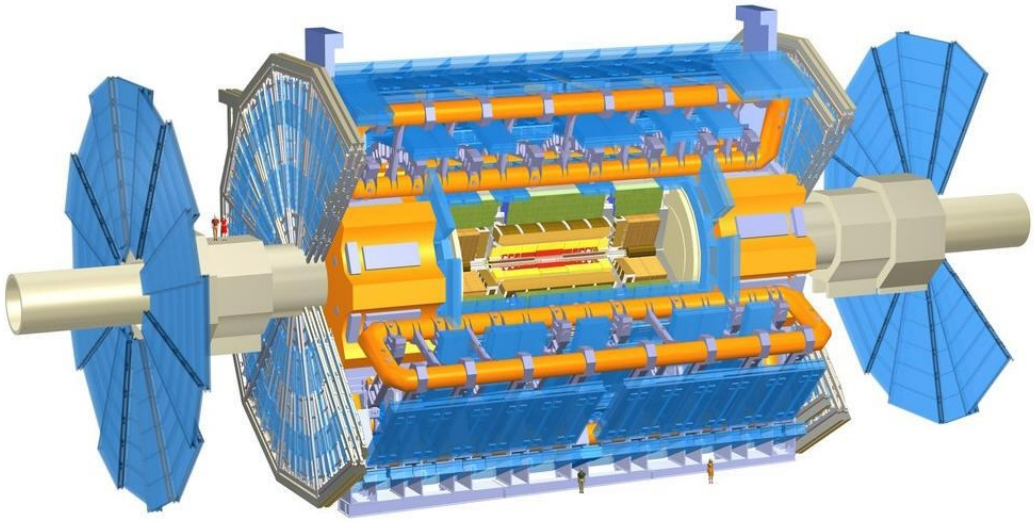

Figure 3.3: A cut-away view of the full ATLAS detector, showing the inner detector, calorimeters, toriod magnet cylinders and the muon spectrometer.

ATLAS Geometry

ATLAS uses a right-handed Cartesian coordinate system, with the x-axis pointing towards the centre of the LHC ring, the y-axis pointing upwards to the surface and the z-axis running parallel to the beampipe. A polar angle coordinate system is also used to simplify descriptions of interactions inside the detector; the polar angle θ is the angle from the beam axis in the $y-z$ plane, and the azimuthal angle ϕ is the angle around the beam axis in the $x-y$ plane. Figure 3.4 gives a diagrammatic representation of this system.

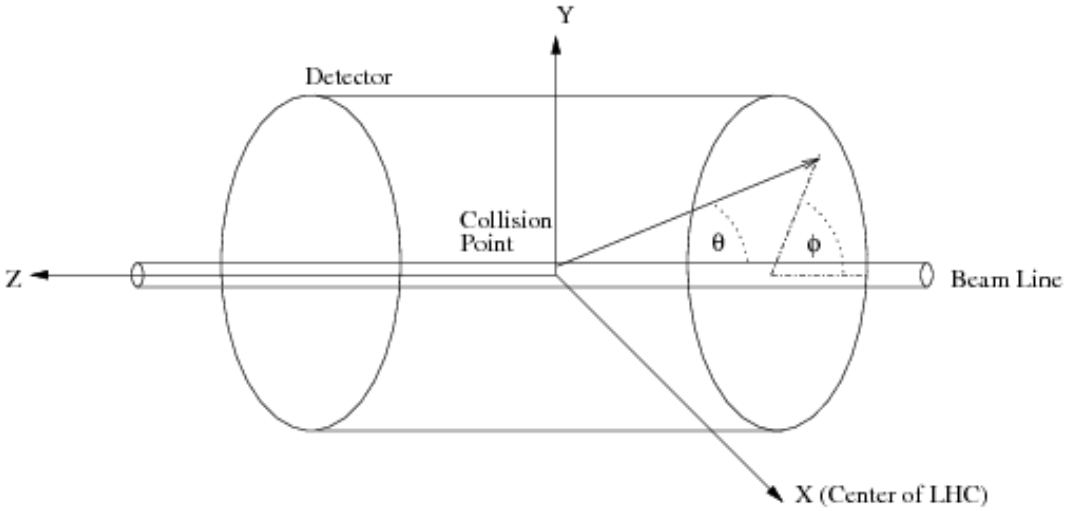

Figure 3.4: The ATLAS coordinate system, showing both the Cartesian xyz system and the polar and azimuthal angles, θ and ϕ. [43]

Given that the LHC is a proton-proton collider, the hard scatter is actually between partons inside the protons, which carry some unknown fraction of the proton momentum. As the protons are moving in only the z-direction, this component is impossible to determine - however the x- and y-components may be measured, and it is convenient to define the tranverse momentum of an object as:

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{3.1}$$

In addition, to get to the centre-of-mass frame, a Lorentz boost in the z-direction is required, so Lorentz invariant quantities are preferred. The rapidity, y, is useful as an angular variable instead of θ as differences in y are invariant under such Lorentz boosts - the rapidity is defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{3.2}$$

However, as p_Z is unknown, the pseudo-rapidity η is used instead of rapidity. This is an approximation of rapidity for a massless particle, but for stable, high-p_T particles, this distinction is negligible. Again, this variable is invariant under Lorentz boosts, and its non-linear nature is useful for describing the forward regions of the detector. It is defined as:

$$\eta = \ln\left(\tan(\theta/2)\right). \tag{3.3}$$

For reference, $\eta = 0$ is directly down the centre of the ATLAS barrel, with $\eta = 1.5$ sitting approximately where the barrel and endcap meet, and $\eta > 5$ basically down the beampipe. The distance between two physics objects in the $\eta - \phi$ plane is then given by ΔR, a Lorentz invariant geometric quantity. ΔR is used for jet formation and resolving overlapping objects in the detector, and is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.4}$$

Another important quantity is the impact parameter, defined as the distance between a primary vertex and an associated secondary vertex . This can be measured in the transverse plane (denoted as d_0) or in the longitudinal plane (denoted as z_0). The better the position of the interaction vertices can be resolved, the more precisely the impact parameter can be measured - this is a driving factor in the development and improvement of the pixel detector. Primary vertices are situated in the beam pipe, and are the underlying proton-proton collision points, while second vertices are caused by the decay of (relatively) long-lived particles, such as B-mesons, that were initially created at the primary vertex. Typically, secondary vertices are also within the beam pipe, though closer to the innermost layer of the detector and distinct from the primary vertex, though if the decaying particle is particularly long-lived, it may occur within the detector itself.

ATLAS Inner Detector

The inner detector (ID) is the closest component to the interaction point, and is designed to track charged particles for momentum measurement and vertex reconstruction. There are three parts to the ID; the pixel tracker, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). These three sub-detectors combine to track particles with $p_T > 0.5$ GeV at angles up to $|\eta| = 2.5$. During the long shut-down in 2013 and 2014, an additional sub-system was added to the pixel tracker, very close to the interaction point, called the Insertable B-Layer (IBL) [44]. The IBL was active during most of the 2015 data-taking, contributing to 87% of the 3.9 fb^{-1} luminosity recorded [42], and through the entire 2016 run. Taking all sub-detectors into account, the ID gives ATLAS a momentum resolution of between 3.8% and 11% for a 100 GeV p_T track, depending on the η of the track, and

between 1.4% and 8% for a 10 GeV p_T track. Each layer in the ID is numbered for identification purposes - layer 0 is closest to the beampipe (not including the IBL), with layer 6 in the SCT being the final silicon tracking layer.

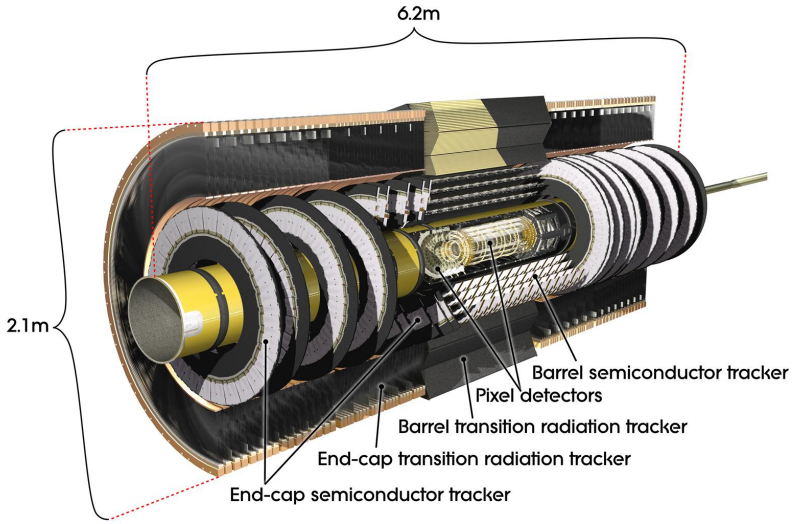

Figure 3.5: A schematic of the inner detector, with each sub-detector labelled. [43]

The Pixel Detector

The pixel detector is a cylindrical sub-detector 48.4 cm in diameter and 6.2 m long, consisting of three concentric layers of silicon pixel strips, situated 50.5 mm, 88.5 mm and 122.5 mm from the beam pipe. Three layers of discs cover each end, providing coverage up to $|\eta| = 2.5$. To cope with the large number of interactions in the LHC, and to provide precise vertex reconstruction and momentum measurement, the pixel detector has the highest granularity of any of the sub-detectors, with 1750 sensor modules connected to 80 million 50×400 μm^2 semiconductor silicon pixels. This gives it a track resolution of 8-14 μm, depending on the track η.

The IBL was added to the pixel detector during the long shutdown between 2012 and 2015. It adds a fourth layer to the pixel detector, closer to the interaction region, without requiring large modifications to the larger system. This new system is mostly used for vertex-finding, b-jet tagging, and fake track reduction at high instantaneous luminosity, as well as providing redundancy to the pixel detector in case of failures or radiation damage. The IBL consists of 14 staves, positioned lengthways along the detector, with 32 front-end chips per stave. The mean radius of the system is 33 mm, with an inner

radius of 31 mm and an outer radius of 38.2 mm [44]. The IBL helps improve the impact parameter resolution of the ID by as much as 50%, in 2015, compared to the 2012 data taken without the IBL [45].

The Semiconductor Tracker

The second subdetector in the ID, the SCT surrounds the pixel detector at a distance of about 300 mm from the beam pipe. Four barrel layers and two endcap layers per side combine to give coverage in the range $|\eta| < 2.5$. Designed to measure charged particle momentum and impact parameter, the SCT has 4088 silicon strip modules with 6 million silicon wafer strips, each just 80 μm wide. The extra distance from the beam pipe compared to the pixel detector allows for a coarser granularity, leading to a position resolution in ϕ of ≈ 16 μm. The next chapter of this thesis details a study performed on the SCT to measure the Lorentz angle of this system, an important quantity for calibration and tracking in the detector.

The Transition Radiation Tracker

The outermost layer of the ID, the TRT forgoes silicon chips and instead uses gaseous straw tube elements and transition radiation materials to provide particle identification capabilities as well as momentum measurement. Each of the $\sim 300,000$ straw tubes is 4 mm in diameter and contains a mixture of xenon, carbon dioxide and oxygen, with a gold-plated tungsten wire running through the centre. A charged particle moving through a straw will ionise the gas; the ionised electrons drift towards the central wire, registering a hit. A consecutive series of hits constitute a track, the curvature of which determines the momentum of the particle. The TRT provides momentum measurement in the range $|\eta| < 2.0$, combining both the barrel and endcaps.

As a particle traverses the TRT, it encounters regions with different dielectric constants - at each boundary, it will radiate energy proportional to the Lorentz factor $\gamma = E/mc^2$. An electron will radiate at X-ray frequencies at energies of 1 GeV or above, whereas for heavier charged particles such as charged pions $\pi^{\pm}$, the X-ray threshold is at approximately 200 GeV. This helps identify electrons without needing full reconstruction - any hit recorded in the X-ray frequency range is only likely to be from electrons, and thus the presence of such a hit can be used to discriminate against pions [46].

ATLAS Calorimetry

The ATLAS calorimeter is split into two sub-detectors, the electromagnetic and the hadronic calorimeters. These calorimeters measure the energy deposits of particles, and also limit the number of 'punch-through' particles into the muon spectrometer, by using alternating layers of active and absorbing

material. The absorbing materials cause particle energy loss through particle showers, and the energy loss is measured in the active layers. Both parts are symmetric in ϕ, and provide combined coverage in the range $|\eta| < 4.9$, though the precision of energy measurement drops above $|\eta| < 2.5$. A large range is needed in order to accurately measure missing transverse energy, as particles with significant transverse momentum may also have large pseudo-rapidity, especially in very high energy events.

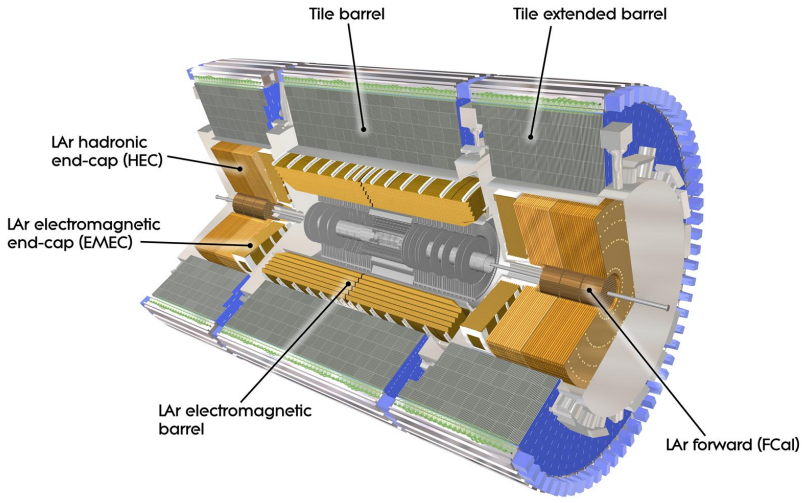

Figure 3.6: Schematic showing the ATLAS calorimeter system. [43]

Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) measures the energy of electrons and photons; muons leave a track but travel straight through the calorimeter. There are three sections, a barrel (covering $|\eta| < 1.475$) and two endcap calorimeters (covering $1.375 < |\eta| < 3.2$), all three of which use liquid argon (LAr) as their active medium and lead as the absorbing medium. Particle showers in the argon liberate electrons that are collected and recorded - this is repeated after each absorption layer, measuring the energy loss of the particles. The barrel section is more than 22 radiation lengths deep, where a radiation length is the distance travelled by a particle before it loses all but $1/e$ of its energy in a given medium; this corresponds to the ECAL fully absorbing electrons and photons with energy less than 50 GeV. Using the calorimeter alone, ATLAS has an energy resolution of 1-1.5% for 100 GeV photons, and 1.3-2.3% for 50 GeV electrons.

Its important to note that there are crack regions in the ECAL at the point where the endcap sections meet the barrel section (approximately $|\eta| = 1.5$), worsening the energy resolution in that region

compared to the rest of the ECAL.

Hadronic Calorimeter

Surrounding the ECAL is the hadronic calorimeter (HCAL), a tile calorimeter that measures the energy of hadronic particles such as protons and pions. Similar to the ECAL, the HCAL has a barrel region and two endcap regions, with the addition of a forward calorimeter to deal with the flux of low transverse momentum particles. The barrel section of the HCAL extends to $|\eta| < 1.0$, with two extended barrel sections covering between $0.8 < |\eta| < 1.7$, and uses concentric layers of steel as the absorbing material and scintillating tiles as the active medium. The scintillation light produced by a traversing particle is collected by wavelength shifting fibres that lead to photomultiplier tubes.

The encap sections of the HCAL use LAr, same as the ECAL, only with tungsten instead of lead as the absorbing medium. The two wheels extend the HCAL coverage between $1.5 < |\eta| < 3.2$. This range is further extended out to $|\eta| < 4.9$ by a forward calorimeter (FCAL) that also uses LAr as the active medium, with one EM layer using copper as the absorbing material and two hadronic layers using tungsten.

Combining the barrel and endcaps, the HCAL has an energy resolution of 2% for 1 TeV jets, and an uncertainty of $\sigma \approx 20$ GeV on the missing energy calculation at total transverse energies of $\Sigma E_T = 2$ TeV.

Muon Spectrometer

The largest and outermost part of the ATLAS detector, the muon spectrometer (MS) surrounds the calorimeters and measures muon momentum as they pass through the detector. A toroidal magnet, varying between 0.5 and 1 T in field strength, bends the path of the charged muons, the curvature of which is proportional to the momentum of the particle. The presence of such a track in the MS is a major identifier of a muon compared to, for example, a high-energy kaon or pion.

The MS is unique in that it has different components for fast online triggering and slower, more precise offline reconstruction. The precision tracking is done using Monitored Drift Tubes (MDT) in the region $|\eta| < 2.0$, and Cathode Strip Chambers (CSC) in the forward region $2.0 < |\eta| < 2.7$. Both are drift chambers - the MDTs have a single central wire while the CSCs are multiwire and faster; the higher speed of the CSCs is required by virtue of the higher event rate in the forward direction. The drift time of the chambers is approximately 100 ns, too slow to be used for hardware-level triggering given the LHC bunch spacing of 25 ns. Thus two faster systems are used, the Resistance Plate Chambers (RPC, covering $|\eta| < 1.05$) and the Thin Gap Chambers (TGC, covering $1.05 < |\eta| < 2.7$). The

RPCs use parallel electrode plates instead of wires, and the TGCs are very fast multiwire proportional chambers - these technologies were chosen as they give a narrow enough time resolution to identify the exact beam crossing a muon comes from, as well as being able to pass the tracking information to the triggering system $O(10$ ns) after a particle passes through.

Combining all these subsystems gives the MS a muon momentum resolution of 3.9% to 6.4% for a 10 GeV p_T muon, depending on its η, and 10.5% to 4.6% for a 1000 GeV p_T muon - interestingly, the momentum resolution is better in the endcaps than the barrel for large-p_T muons.

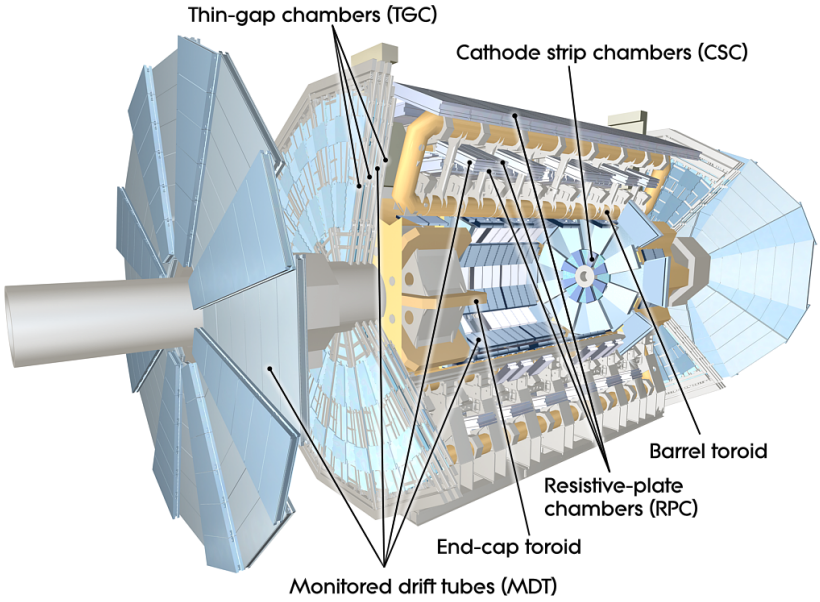

Figure 3.7: A cut-away view of the ATLAS muon spectrometer.

ATLAS Trigger System

The bunch crossing frequency of the LHC is 40 MHz, with just 25 ns in between collisions. The rate at which ATLAS can record data is 1 kHz - therefore the vast majority of events must be discarded in order to feasibly store the data. A multi-level trigger system is used to do this, quickly identifying and rejecting uninteresting events while passing the interesting events up the chain to be stored and processed. In 2011 and 2012, ATLAS ran a three-level system: a Level 1 (L1) hardware based trigger, a Level 2 (L2) software based trigger, and an Event Filter (EF) that used the full reconstruction algorithm to select events. For Run 2 this system was adapted into a two-level system, with a L1 hardware trigger and a High Level Trigger (HLT), the advantages of which include reducing

bandwidth bottlenecks arising from the network structure, and being able to use faster algorithms closer to reconstruction level.

Level 1 Trigger

The L1 trigger uses the raw information from the MS and calorimeters (reconstructing the inner detector tracks would cause the system to be too slow) to identify high-E_T physics objects. Regions of Interest (RoIs) are defined around these physics objects, and are used as a seed for the software-based triggers. During Run 1 the maximum rate of the L1 trigger was 75 kHz - upgrades made to the system during the long shutdown in 2013 and 2014 improved this maximum rate to 100 kHz. As it takes 2.5 μs for the digital signal to reach the front-end electronics, the L1 trigger uses custom fast electronics - the event data from the sub-detectors are stored in memory buffers before being processed and sent to the HLT.

High Level Trigger

In Run 1, the HLT was composed of the L2 trigger and the EF. The L2 trigger used the RoI information from the L1 system to seed its own selection. As a lower rate system than the L1, the L2 had access to fine-granularity information from all sub-detectors including the inner detector. Fast algorithms process partial event information within the RoIs to identify physics objects and select events based on their kinematics. The Event Filter then takes those selected events and uses the full reconstructed event to determine whether to save the event, reducing the final output rate to 400 Hz.

In Run 2, this dual-level trigger system was combined together into a single software-based decision farm. This was needed due to the changing run conditions, moving from $\sqrt{s} = 8$ TeV to 13 TeV; pileup increases from an average of 21 interactions per bunch crossing to an estimated 55, event sizes on disk increase, and the higher luminosity and smaller bunch spacing requires an increase in overall trigger rates. Upgrades to the readout system from the L1 electronics, improved data transfer capacity and new algorithms for building RoIs all contributed to acheiving this new rate.

Trigger Chains

A trigger 'chain' is the list of triggers used during the event selection process. Each chain starts with an L1 trigger and ends with either an EF trigger for a chain used in Run 1, or an HLT trigger for Run 2 chains. A simple example would be a single jet trigger chain from 2012: an L1 trigger requires 1 jet with $p_T > 30$ GeV, seeds an L2 trigger with a tighter p_T cut, then the event is passed to the EF with an even tighter selection.

$$\text{L1\_J30} \rightarrow \text{L2\_j50} \rightarrow \text{EF\_j55}$$

This also demonstrates the acronym scheme for physics objects - for example, a Run 2 missing energy trigger with an energy threshold of 150 GeV would be named HLT_xe150, or a single photon trigger with a transverse momentum threshold of 120 GeV would be called HLT_g120_loose, where the _loose refers to the isolation criterion, dictating how separated in ΔR an object has to be from other objects in the detector in order to qualify for the trigger requirement.

Chapter 4

The SCT Lorentz Angle

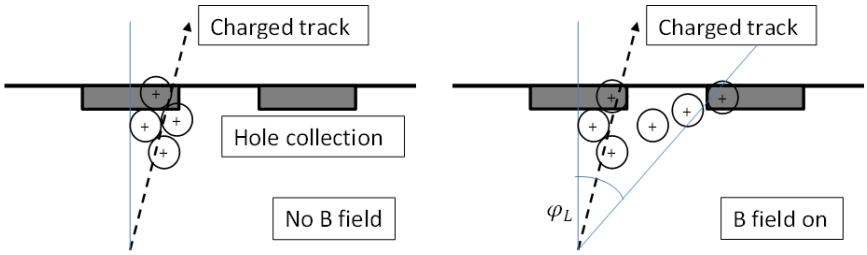

Figure 4.1: Representation of the effect of the magnetic field on a charged particle track reconstructed in the SCT, where ϕ_L is the Lorentz angle of the drifting charge carriers [47]. The left diagram shows the behaviour under no magnetic field - the holes collect in the same module the track passes through. On the right, the magnetic field is turned on, and the holes are deflected by the angle ϕ_L, with some ending up in the neighbouring module - the track is smeared!

The ATLAS SCT is a silicon strip detector designed to measure charged particle trajectories, described more fully in the previous chapter. Charged particles crossing the silicon strips create electron-hole pairs, the holes of which drift to electrodes, guided by an electric field called the bias voltage, and are collected as a signal in a strip module. During the offline event reconstruction, the electronic signals produced by these holes are combined to make a particle track through the SCT. As the SCT is inside a 2T magnetic field, the drift path of the holes is deflected away from the original particle track, as demonstrated in Figure 4.1. The angle of this deflection is called the Lorentz angle. In general the Lorentz angle is quite small and holes will be collected at the same strip module that the charged particle passed through, but for tracks with large angles with respect to the magnetic field direction, the deflection may cause a small number of holes to be collected in an adjacent module. This introduces some level of uncertainty in the track position, when the signals from the strip modules are combined to reconstruct the path of the particle, and understanding the behaviour of the Lorentz angle is key

to minimising this uncertainty.

This chapter provides an update on the Lorentz angle measurements from proton-proton collisions and incident cosmic rays. With the increase of beam energy from 8 TeV to 13 TeV and the introduction of the IBL, the conditions in the SCT have changed in 2015 with respect to 2012, and may affect the Lorentz angle. The SCT Lorentz angle is estimated to be about 4 degrees, corresponding to a shift in the collected hole position of approximately 20 μm across the 80 μm strip pitch and 285 μm sensor width. The Lorentz angle depends upon the detector conditions, such as the bias voltage and operating temperature, as well as the external magnetic field.

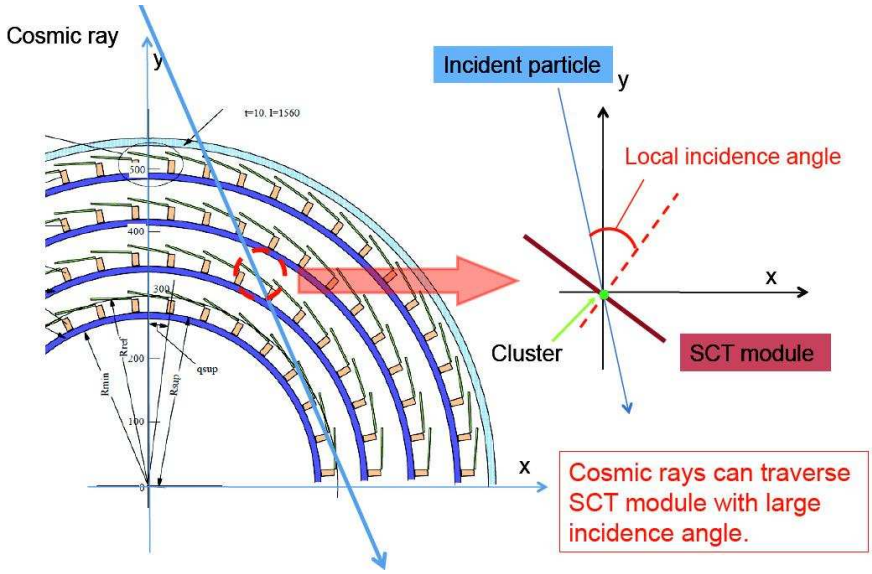

Figure 4.2: Cut-away view of the SCT barrel section, and a diagrammatic representation of a particle track incident upon an SCT module [48].

This chapter is organised as follows. Section 4.1 describes the theoretical prediction of the Lorentz angle, using two different models. In section 4.2, the procedure for extracting the Lorentz angle is described, including the fitting function used. Section 4.3 uses collision data taken from a special $\sqrt{s} = 13$ TeV commisioning run with a low number of interactions per bunch crossing to measure the Lorentz angle under collision conditions. The calculation of the systematic uncertainties contributing to this measurement is described in section 4.4, with some discussion following in section 4.5.

4.1 Theoretical predictions

The Lorentz angle (ϕ_L) is a function of the charge carrier mobility and the external magnetic field, expressed as

$$\tan\phi_L = \mu_H \cdot B = r \cdot \mu_d \cdot B \tag{4.1}$$

where μ_H is the Hall mobility, μ_d is the drift mobility and r represents the Hall scattering factor (around 0.72 for holes in silicon). The drift mobility is given by

$$\mu_d = \frac{\mu^0}{(1 + (\frac{\mu^0 \cdot E}{\nu_{\mathrm{sat}}})^\beta)^{1/\beta}} \tag{4.2}$$

using the electric field E in the sensor, the saturation velocity ν_{sat} and a parameter β that expresses the transition between ohmic behaviour and saturation velocity. The low field mobility μ^0 is given by ν_{sat}/E_c, where E_c is the electric field scale at which the transition from the linear regime to saturation occurs. Two models have been used to find the expected values for the SCT drift mobility: the Jacoboni-Canali (JC) parameterisation [49, 50] and the Becker-Fretwurst-Klanner (BFK) model [51], both of which are described below. The main difference between the two models is that the BFK model distinguishes between different crystal orientations, denoted by their Miller indices, predicting different Lorentz angle values for each orientation. This is pertinent to the SCT given that while the majority of the strip modules were produced in the <111> crystal orientation, there are some that are in the <100> crystal orientation. The JC model does not make this distinction, potentially giving incorrect predictions for the <100> modules, which could be corrected by the BFK model.

The electric field in the SCT is close to what is expected for a flat diode, with the exception of the hole collection regions in the strip modules themselves. It is therefore given by:

$$E_Z = \frac{V_B + V_D}{d} - 2\frac{V_D}{d^2}z \tag{4.3}$$

where V_B is the bias voltage (150 V), V_D is the depletion voltage (70 V), d is the sensor thickness ($285\mu m$), and z is the position at which the field is being evaluated. For the SCT, the Lorentz angle is estimated in the center of the sensor at $z = d/2$, therefore $E = 5263$ V/cm.

Jacoboni-Canali model

The JC model [49, 50] for holes uses the following parameterisation:

$$\nu_{\text{sat}} = 1.62 \cdot 10^8 \cdot T^{-0.52} \text{ cm/s} \tag{4.4}$$

$$E_c = 1.24 \cdot T^{1.68} \text{ V/cm} \tag{4.5}$$

$$\beta = 0.46 \cdot T^{0.17} \tag{4.6}$$

where T is the temperature in kelvin. The expected values of μ_d and the Lorentz angle ϕ_L for each layer of the SCT in the barrel are stated in Table 4.1. Systematic uncertainties are dominated by the drift velocity and the non-uniformity of the electric field, both of which are estimated to be at the level of 5% uncertainty. Small contributions from the temperature T and magnetic field B are also included, estimated at 0.5% and 0.75% uncertainty respectively. Other sources of uncertainty, such as the bias voltage and impurities in the strips, are predicted to be negligible.

Barrel	T [K]	μ_d $[cm^2/(V \cdot s)]$	ϕ_L [deg]
3	279.5	446±34	3.67±0.26
4	278.5	449±34	3.70±0.26
5	279.7	445±34	3.67±0.26
6	279.6	445±33	3.67±0.25

Table 4.1: Expected drift mobility and Lorentz angle for the four SCT barrel layers in the JC model, with estimated systematic uncertainties. Barrel layers 0, 1 and 2 are in the pixel detector and are excluded from this study.

Becker-Fretwurst-Klanner model

The BFK model [51], first adapted for the ATLAS SCT in the 2013 Lorentz angle measurement [52], estimates the drift mobility by starting from measured parameters at 300 K, shown in table 4.2, and using the exponential law in Eq. 4.7 to evaluate the expected parameters at a given temperature T.

$$P_i(T) = P_i(300K) \cdot \left(\frac{T}{300}\right)^{\alpha_i} \tag{4.7}$$

where $P_i(300K)$ refers to one of the parameters stated in table 4.2 at the measured temperature of 300 K. The values for the exponent α_i are also given by the model, and are stated in table 4.3. Expected Lorentz angle values for this model are stated in Table 4.4.

μ_0 [cm^2/(V·s)]	ν_{sat} [cm/s]	$\beta^{<100>}$	$\beta^{<111>}$
474±10	(0.940±0.027)·10^7	1.181 ± 0.003	0.924 ± 0.002

Table 4.2: Measured parameters at 300 K in the BFK model, with estimated systematic uncertainties included.

	μ_0	ν^{sat}	$\beta^{<100>}$	$\beta^{<111>}$
α_i	-2.619±0.007	-0.226±0.002	0.644 ± 0.003	0.550 ± 0.002

Table 4.3: Temperature parameters α_i for the BFK model parameters, with estimated systematic uncertainties included.

Barrel	T [K]	$\phi_L^{<100>}$	$\phi_L^{<111>}$ [deg]
3	279.5	3.80±0.25	3.33±0.22
4	278.5	3.82±0.25	3.35±0.22
5	279.7	3.79±0.25	3.33±0.22
6	279.6	3.80±0.25	3.33±0.22

Table 4.4: Expected Lorentz angle values using the BFK model, in both the <111> and <100> module crystal orientations. Quoted errors are from systematic uncertainties only.

4.2 Lorentz Angle Extraction

The two important quantities in determining the Lorentz angle in the ATLAS detector are the cluster width and the incident angle. A cluster is a collection of hits in adjacent SCT strips, the size or width of which is the number of strips within it. The incident angle ϕ is measured in the plane defined by the normal to the surface of the strip and the axis in the strip direction. Without a magnetic field present, the minimum cluster width occurs at an incident angle perpendicular to the plane of the SCT module. When a magnetic field is applied, a Lorentz force is generated in the strips and the minimum cluster width instead occurs at the Lorentz angle.

To extract the Lorentz angle, profile plots of the average cluster width versus the incident angle are used. The data are fitted using the function:

$$\begin{aligned} f(\phi) &= (a|\tan\phi - \tan\phi_L| + b) \otimes \text{Gaussian}(\phi) \\ &= \int_{-\infty}^{\infty} (a|\tan\phi' - \tan\phi_L| + b) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\phi'-\phi)^2}{2\sigma^2}\right) d\phi' \end{aligned} \tag{4.8}$$

where ϕ_L is the Lorentz angle, a is the slope parameter, b is the minimum cluster width and σ represents

the smearing effects such as charge carrier diffusion [48,52]. All four parameters are left free and fitted from data.

4.3 Lorentz Angle in Collision Data

In 2015 the LHC was primarily run at $\sqrt{s} = 13$ TeV with two different bunch spacings, 50 ns and 25 ns, with a limited number of special runs commisioned for testing and low-pileup physics. Pileup (or μ) refers to the number of interactions per bunch crossing - higher pileup means more interactions, with more primary vertices and more physics objects. Whilst the pileup is expected to not have any effect on the Lorentz angle, a low-pileup run ($\mu = 0.1$, compared to the 2012 average $\mu = 20.7$) from prior to the start of 50 ns running has been used here, for its large statistics and cleanliness of tracks (each track is distinct with a clear separation between neighbouring tracks - the same may not be true in high-μ events). A tight track selection is applied, the details of which are below:

- Track $p_T > 500$ MeV
- > 6 hits in the SCT
- > 1 pixel hits
- Track $|d_0| < 1$mm

where the track $|d_0|$ is the impact parameter, defined as the shortest perpendicular distance between the primary vertex and the projection of the secondary vertex onto the beam axis. The impact parameter is used to discriminate against particles coming from sources other than a bunch crossing interaction, such as cosmic muons or so-called 'beam splashes' where a proton bunch hits a collimator in the beam pipe downstream of the detector and sends a spray of particles into the detector at an unexpected angle. In addition to these cuts, only negatively charged tracks are selected. This is due to the tilt angle the silicon sensors are positioned at, with respect to the beam axis, which results in only the negatively charged tracks being able to reach the Lorentz angle as the positively charged tracks are bent in the wrong direction by the magnetic field.

Figure 4.3 shows the incident angle versus average cluster width for this run, for both the <111> and <100> modules. The data are fitted using the function shown in Eq. 4.8. Table 4.5 contains the values for the Lorentz angle in each barrel layer and module orientation, as obtained by the fit. Systematic uncertainties are included, the evaluation of which follows the procedure detailed in Section 4.4.

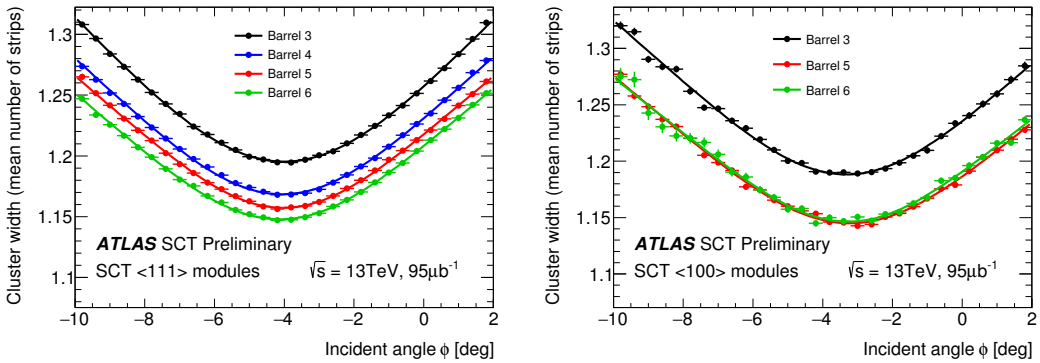

Figure 4.3: Incident angle vs average cluster width for the <111> (left) and <100> (right) modules. The curve overlaid onto the data shows the result of the fit - the minimum of each curve is the Lorentz angle in that layer.

Barrel	T [K]	ϕ_L(<111>) [degrees]	ϕ_L(<100>) [degrees]
3	279.5	4.02 ± 0.01	3.33 ± 0.03
4	278.5	4.07 ± 0.01	-
5	279.7	4.04 ± 0.01	3.23 ± 0.02
6	279.6	4.13 ± 0.01	3.28 ± 0.03

Table 4.5: Lorentz angle values obtained from the fit to data, for the <111> and <100> modules in each barrel layer. Only statistical errors are shown.

4.4 Systematic Uncertainty Evaluation

The systematic uncertainty is dominated by the uncertainty caused by the Lorentz angle extraction procedure itself, due to the assumptions made in the fitting function. This is assessed in two ways: first, by fitting cosmic ray data taken with no magnetic field and assessing the systematic deflection away from $\phi_L = 0$ - the Lorentz angle depends upon the magnetic field strength, so a non-zero value measured with no B-field is a systematic bias in the fit. Second, by fitting the data with an independent asymmetric function described in Eq. 4.9 rather than the standard symmetric function, as the data is clearly asymmetric about the minimum value. Any difference between the standard symmetric fit and the asymmetric fit is then taken as a contribution to the systematic uncertainty due to this asymmetry about ϕ_L.

$$f(\theta) = \begin{cases} (a_1 \cdot |\tan\phi - \tan\phi_L| + b) \otimes \text{Gauss} & (\phi < \phi_L) \\ (a_2 \cdot |\tan\phi - \tan\phi_L| + b) \otimes \text{Gauss} & (\phi \geq \phi_L) \end{cases} \tag{4.9}$$

The cosmic ray data were taken from a single run in March where the magnetic field was turned off, but both the muon spectrometer and the inner detector were running normally. A very loose track selection is applied, requiring only > 4 SCT hits and track $|d_0| < 500$ mm. The standard symmetric function is then fitted to the data to extract the Lorentz angle, the result of which is shown in figure 4.4.

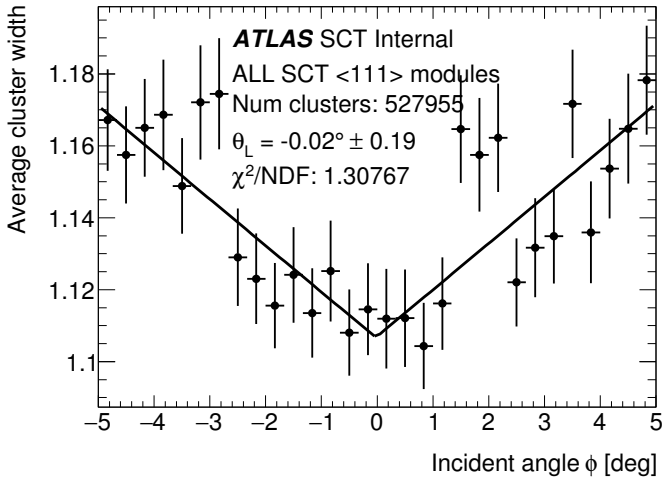

Figure 4.4: Cosmic ray data taken with no magnetic field, <111> modules only, with all barrel layers combined to maximise statistics. The curve shown is the result of the symmetric fit.

The Lorentz angle value from these data is $0.02 \pm 0.19°$. This is consistent with 0°, and also with the cosmic ray measurement from 2011, $0.07 \pm 0.05°$ [48]; given the limiting number of statistics in the cosmic ray data used for this new measurement, the maximum possible value of the 2011 measurement, 0.12°, has been used as the value of the systematic uncertainty contribution.

Figure 4.5 shows the asymmetric function fitted to the collision data, for both module orientations. Extracting the Lorentz angle from these fits gives the values shown in Table 4.6. The difference between the asymmetric values and the symmetric values is small - the large statistics results in a more stable fit, leading to a small systematic uncertainty from the Lorentz angle extraction.

The average difference $< \Delta >$ between the asymmetric and symmetric fit functions is 0.05 for both module orientations. The overall systematic uncertainty is calculated by summing in quadrature the cosmic ray measurement and $< \Delta >$, as shown in equation 4.10.

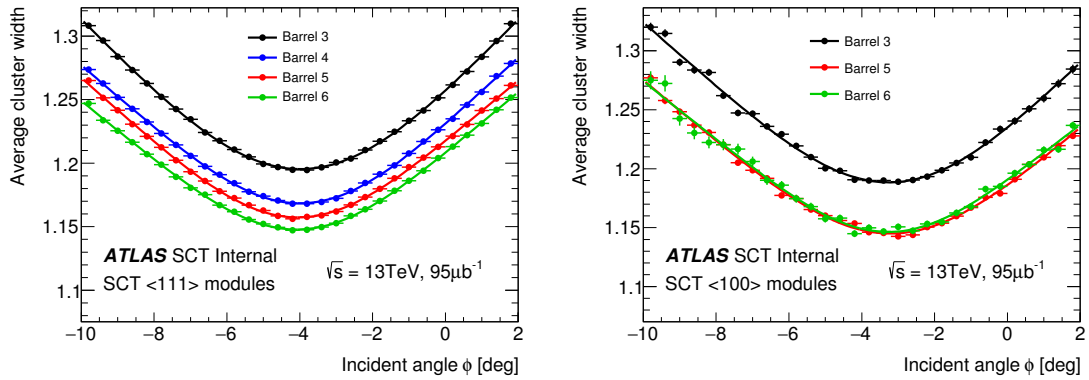

Figure 4.5: Incident angle vs average cluster width for the <111> (left) and <100> (right) modules, fitted using the asymmetric function

	<111> [degrees]			<100> [degrees]		
Barrel	$\phi_L^{asymmetric}$	$\phi_L^{symmetric}$	Δ	$\phi_L^{asymmetric}$	$\phi_L^{symmetric}$	Δ
3	3.98 ± 0.03	4.02 ± 0.01	0.03	3.20 ± 0.12	3.33 ± 0.03	0.13
4	3.98 ± 0.01	4.07 ± 0.01	0.09	-	-	-
5	4.07 ± 0.01	4.04 ± 0.01	0.03	3.20 ± 0.06	3.23 ± 0.02	0.03
6	4.08 ± 0.02	4.13 ± 0.01	0.05	3.28 ± 0.08	3.28 ± 0.03	0.00

Table 4.6: Lorentz angle values for asymmetric and symmetric fits, in both <111> and <100> modules, with the difference between the two fit functions shown as Δ.

$$\begin{aligned} \epsilon &= \sqrt{\text{cosmics}^2 + <\Delta>^2} \\ &= \sqrt{0.12^2 + 0.05^2} \\ &= 0.13 \end{aligned} \tag{4.10}$$

4.5 Summary

Figure 4.6 shows a summary of all the Lorentz angle values, for both theoretical models and the data in both <111> and <100> modules, with systematic uncertainties included. The Lorentz angle in data is consistent across all layers, however it does not agree well with either theoretical model.

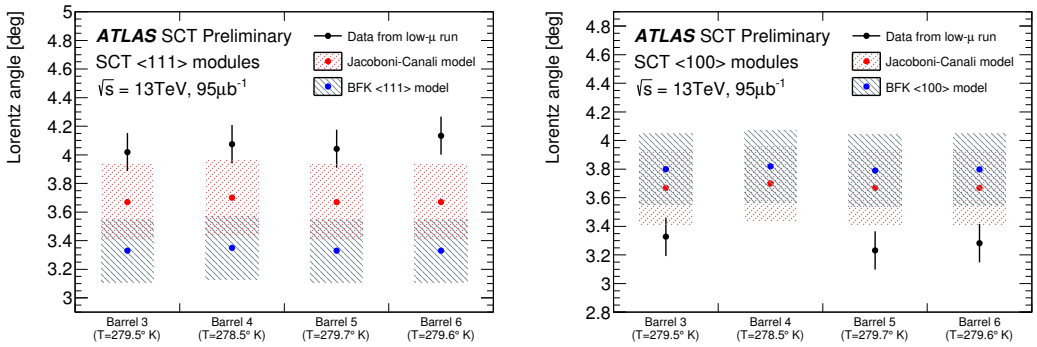

Figure 4.6: Summaries of the Lorentz angle values for both the Jacoboni-Canali and BFK models, and the measured values from data in both <111> (left) and <100> (right) modules.

With the insertion of the IBL, the temperature of the SCT has changed, and the Lorentz angle could have potentially changed as as a result. A dramatic increase in statistics compared to previous measurements allows for a more precise measurement of the Lorentz angle in both <111> and <100> modules.

Comparing the results from data to the two theoretical models studied here, the Jacoboni-Canali model and the BFK model, shows that the data do not agree well with the predictions. There are consistent discrepancies across both crystal orientations and all barrel layers - these discrepancies are larger for the BFK model, which provides a distinction between the two module orientations and therefore should provide more accurate predictions. Disagreements with the models have been presented in previous measurements, ruling out the temperature as a cause; this is yet to be understood.

The Lorentz angle values calculated here are fed back into the detector response simulations, to make sure the modelling of the detector is consistent under the current running conditions.

Chapter 5

Rare B_s^0 Decays as a New Physics Probe

5.1 Introduction

Flavour changing neutral currents (FCNC) are heavily suppressed in the SM, as previously discussed in Chapter 2. This makes them excellent candidates for indirectly probing new physics effects, as small deviations caused by new particles are no longer swamped by the much larger SM backgrounds. Of particular interest are the rare B decays, $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, FCNC processes that have very small SM branching fractions, predicted to be $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.65 \pm 0.23) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10}$ [53], due to additional helicity considerations that further suppress these processes. BSM theories such as the MSSM, two-Higgs-Doublet models, and Minimal Flavour Violation models [54–56] predict new interactions that cause deviations away from the SM in the $B_s^0 \to \mu^+\mu^-$ branching fraction. The CMS and LHCb collaborations reported the first observation of this process using $\sqrt{s} = 7$ and 8 TeV LHC data, with measured branching fractions of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \left(2.8^{+0.7}_{-0.6}\right) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = \left(3.9^{+1.6}_{-1.4}\right) \times 10^{-10}$ [57]. This observation has since been confirmed by LHCb alone, using $\sqrt{s} = 13$ TeV data collected in 2015 and 2016, with a measured branching ratio of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.8 \pm 0.6) \times 10^{-9}$, in agreement with the SM prediction [35].

Because the total cross-section for $b\bar{b}$ collisions at the LHC is not well known, the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions are measured relative to the normalisation decay $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, which has a known branching fraction, $\mathcal{B}(B^+ \to J/\psi\, K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)$. By measuring a relative branching fraction, rather than the absolute, the uncertainties related to this unknown $b\bar{b}$ total cross-section are cancelled out. The procedure needs to be corrected firstly for the ratio of the different probabilities of the b quark to hadronise with an u or a s (d) quark; this correction enters as the relative b-quark hadronisation probability f_u/f_s (f_u/f_d), though here we can make the assump-

tion that $f_u = f_d$. Secondly, the different selection efficiencies $\varepsilon_{\mu^+\mu^-}$ and $\varepsilon_{J/\psi K^+}$ must be taken into account. In the simplest form, the $B^0_{(s)} \to \mu\mu$ branching fraction can be extracted as:

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = \frac{N_{d(s)}}{\varepsilon_{\mu^+\mu^-}} \times \left[\mathcal{B}(B^+ \to J/\psi\, K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)\right] \frac{\varepsilon_{J/\psi K^+}}{N_{J/\psi K^+}} \times \frac{f_u}{f_{d(s)}}\,,$$

where $N_{d(s)}$ is the $B^0_{(s)} \to \mu\mu$ signal yield, and $N_{J/\psi K^+}$ is the $B^+ \to J/\psi\, K^+$ normalisation yield [58]. For the ATLAS study, a modified formula is used to normalise independently samples of events collected in different data-taking periods and with different trigger selections:

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = N_{d(s)} \times \left[\mathcal{B}(B^+ \to J/\psi\, K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)\right] \times \frac{f_u}{f_{d(s)}} \times \frac{1}{\mathcal{D}_{\text{norm}}}\,, \tag{5.1}$$

with

$$\mathcal{D}_{\text{norm}} = \sum_k N^k_{J/\psi K^+} \alpha_k \left(\frac{\varepsilon_{\mu^+\mu^-}}{\varepsilon_{J/\psi\, K^+}}\right)_k\,. \tag{5.2}$$

The denominator $\mathcal{D}_{\text{norm}}$ consists of a sum whose index k runs over each of the data-taking periods and the trigger selections (see Section 5.4). In the sum, the α_k parameter takes into account the different trigger prescale factors and integrated luminosities in the signal and normalisation channels, and the ratio of the efficiencies corrects for reconstruction differences in each data sample k.

The outline for this chapter is as follows: Section 5.2 details the data and MC samples used in this study, and Section 5.3 defines the candidate event selection, including the muon definition. The trigger selection strategy used to maximise the available statistics is explained in Section 5.4 with Section 5.5 detailing the background sources and composition. Section 5.6 goes through the multivariate analysis (MVA) techniques used to separate the signal from the background, and Section 5.7 uses the variables in the MVA to compare data and MC in order to assess any systematic effects from event mis-modelling. The B^+ reference channel yield is extracted in Section 5.8. The efficiency ratio between signal and normalisation channels, $\mathcal{D}_{\text{norm}}$, is calculated in Section 5.9, before the signal yield and $B^0_s \to \mu^+\mu^-$ branching ratio are calculated in Sections 5.10 and 5.11.

My personal contributions towards this result are the study and definition of the trigger categories, in Section 5.4, and the binned B^+ fits and data-MC comparisons, in Section 5.7. Some contribution was also made towards the calculation of the $B^0_s - B^+$ efficiency ratios, as the B^+ data-MC comparisons are used to evaluate the effect of MC reweighting on the efficiency ratios.

5.2 Data and Monte Carlo Samples

This analysis uses the Run 1 data sample, recorded in 2011 and 2012 by ATLAS, from pp collisions at $\sqrt{s} = 7$ TeV and 8 TeV respectively. After data quality requirements are imposed, based on the performance of the muon spectrometer (MS) and inner detector (ID) systems, the total integrated luminosity is 4.9 fb^{-1} for 2011 and 20 fb^{-1} for 2012. The analysis was performed 'blind' - any data with a dimuon invariant mass in the signal region $5166 - 5566$ MeV, or ± 200 MeV around the B_s^0 mass, was omitted from the analysis development, in order to avoid biasing the selection optimisation. Sideband data, where the dimuon invariant mass is outside the signal mass range but still within the acceptance of the analysis ($4766 < m(\mu\mu) < 5966$ MeV) is used for testing the optimisation instead.

Samples of simulated MC events are used for training and validation of the multivariate analyses, for the determination of the efficiency ratios, and for guiding the signal extraction fits. Exclusive MC samples were produced for the signal channels $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$, the normalisation channel $B^+ \to J/\psi\, K^+$ ($J/\psi\mu^+\mu^-$), the $B^+ \to J/\psi\, \pi^+$ channel, and the control channel $B_s^0 \to J/\psi\, \phi$ ($\phi \to K^+K^-$). In addition, background studies employ MC samples of inclusive semileptonic decays $B \to \mu X$, plus samples of $B_s^0 \to K^-\mu^+\nu$, $B^0 \to \pi^-\mu^+\nu$, $\Lambda_b \to p\mu^-\overline{\nu}$, $B^0_{(s)} \to hh'$ decays with $h^{(\prime)}$ being a charged pion or kaon, and inclusive decays $B \to J/\psi X$.

Most of dimuon candidates in the data sample originate from the uncorrelated decays of hadrons produced in the hadronisation of b and $\bar{b}$ quarks. To describe this background, named the 'continuum', a large MC sample was generated by selecting specific topologies that dominate it, considering both the primary decays from b quarks and the secondary decays from c quarks. Independent samples of events with forced semileptonic decays or decays including muons pairs from J/ψ were generated in all combinations, with the total number of events in each sample chosen carefully to reproduce the expected overall composition of oppositely charged muon pairs in data.

The MC samples were generated with Pythia 6 [59] for studies related to data collected in 2011, and with Pythia 8 [60] and EvtGen [61] for the 2012 samples and the development of multivariate classifiers. The ATLAS detector and its response are simulated using Geant4 [62]. Additional pp interactions in the same and nearby bunch crossings (pile-up) are included in the simulation. All simulated samples are reweighted to have the same distribution of the number of primary vertices (PVs) per bunch crossing found in data.

The simulated samples of the exclusive decays considered are adjusted with two-dimensional data-driven weights (DDW) to correct for the differences between simulation and data observed in the transverse B meson momentum (p_T^B) and B meson pseudo-rapidity ($|\eta^B|$) distributions. DDW obtained from $B^+ \to J/\psi\, K^+$ decays are used to correct the simulation samples in the signal and normalisation

channels. DDW obtained from the $B_s^0 \to J/\psi\,\phi$ control channel are found to agree with those from $B^+ \to J/\psi\,K^+$, though there is a weak dependence on the final state multiplicity.

Similarly to the exclusive decays, the large continuum background MC sample is reweighted via DDW obtained from its comparison with the data in the sidebands of the signal region.

5.3 Candidate Event Selection

After offline reconstruction, a preliminary selection is performed on candidates for $B^0 \to \mu^+\mu^-$, $B^+ \to J/\psi\,K^+ \to \mu^+\mu^-K^+$ and $B_s^0 \to J/\psi\phi \to \mu^+\mu^-K^+K^-$ decays. In the ID system, muons are required to have at least one hit (a cluster of energy deposits indicating a charged particle has passed through the sensor) in the pixel detector, five hits in the semiconductor tracker, and six hits in the transition-radiation tracker, if $0.1 < |\eta| < 1.9$. They are also required to be reconstructed in the MS, and to have $|\eta| < 2.5$ and $p_T > 4$ GeV. Kaon candidates have to satisfy similar requirements in the ID, except that at least nine instead of six hits are required in the transition-radiation tracker and a looser requirement of $p_T > 1$ GeV is imposed.

Event properties are computed based on a secondary vertex (the point of decay for the B meson, which has a long enough lifetime to travel a short distance from the primary vertex) fitted to two, three or four tracks, depending on the decay process to be reconstructed. A goodness-of-fit test is performed on this vertex fit, whereby the χ^2 per degree of freedom of the fit is required to be less than six for the B vertex, and less than ten for the $J/\psi \to \mu\mu$ vertex. The conditions $2915 < m(\mu\mu) < 3275$ MeV and $1005 < m(KK) < 1035$ MeV are required on ID track combinations for the $J/\psi \to \mu\mu$ and the $\phi \to KK$ vertices, respectively. In the $B^+ \to J/\psi\,K^+$ fits the reconstructed J/ψ mass is constrained to the world average value [3].

Reconstructed B candidates are required to satisfy $p_T^B > 8.0$ GeV and $|\eta^B| < 2.5$. The dimuon invariant mass for $B^0 \to \mu^+\mu^-$ candidates is calculated using the combined ID and MS information, in order to improve the mass resolution in the end-caps with respect to using ID information only [63].

The invariant mass range considered for the $B^0 \to \mu^+\mu^-$ decay is 4766–5966 MeV in which the 5166–5566 MeV range is defined as the signal region while the low-mass and high-mass regions (4766–5166 and 5566–5966 MeV) are the signal mass sidebands. For the reference channel, the mass range considered is 4930–5630 MeV for $B^+ \to J/\psi\,K^+$ in which the 5180–5380 MeV range is the peak region and the low and high mass ranges either side of this are the sidebands. This analysis is performed 'blind' - data within the $B_s^0 \to \mu^+\mu^-$ signal region are left out of optimisation studies to avoid any bias towards the signal. The $B_s^0 \to \mu^+\mu^-$ sidebands and the $B^+ \to J/\psi\,K^+$ reference channel are used to check how well the MC models the SM background. Once systematic effects and MC modelling

are deemed to be well-understood, the $B_s^0 \to \mu^+\mu^-$ signal region is unblinded and the data within is compared to the simulation predictions.

The coordinates of the PVs are obtained from charged tracks not used in the decay vertices, and are required to be within the interaction volume of the detector. The matching of a B candidate to a PV is made by propagating the associated tracks backwards to the point of closest approach to the collision axis, and choosing the PV with the smallest separation along z. Simulation shows that this method achieves a correct matching probability of better than 99%.

To reduce the large background in the $B^0 \to \mu^+\mu^-$ channel before the optimisation of the signal region, a number of additional selections are applied, which depend on the distance between the primary and secondary vertices, $\overrightarrow{\Delta x}$. L_{xy}, the projection of $\overrightarrow{\Delta x}$ in the direction of the B momentum $\overrightarrow{p_T}^B$, is required to be > 0. The absolute value of the difference in azimuthal angle, $|\alpha_{2\mathrm{D}}|$, between $\overrightarrow{\Delta x}$ and $\overrightarrow{p_T}^B$ is required to be smaller than 1.0 rad. Finally, using the difference in rapidity $\Delta\eta$ between $\overrightarrow{\Delta x}$ and $\overrightarrow{p_T}^B$, the combination $\Delta R = \sqrt{\alpha_{2\mathrm{D}}^2 + \Delta\eta^2}$ is required to be smaller than 1.5. These requirements reduce the background by a factor of 2.5, with a signal efficiency of 95%.

After the preliminary selection, approximately 2.6×10^6 $B^0 \to \mu^+\mu^-$ and 2.3×10^6 $B^+ \to J/\psi\, K^+$ candidates are found in the respective channels.

5.4 Trigger Categories

For data collected during the LHC Run 1, the ATLAS detector used a three-level trigger system, consisting of a hardware-based Level-1 trigger, software-based Level-2 and Event Filter (EF) triggers, as described in Chapter 3.

A dimuon trigger [64] is used to select events. The 2011 data sample contains events seeded by a Level-1 dimuon trigger that required a transverse momentum $p_T > 4$ GeV for both muon candidates. Due to the increased pile-up in 2012 data, this dimuon trigger was prescaled at the beginning of every fill. Prescaling is a process used to reduce the number of triggered events to a level that is manageable by the next level of the trigger chain. Usually it involves only passing one in X events onto the next stage, where X varies depending on the instantaneous luminosity and trigger output rate. The effect of this prescaling can be seen in Figure 5.1, as compared to an unprescaled trigger.

This prescaling reduced the useable data available for measurement, from the 20.3 fb^{-1} total luminosity collected in 2012 down to just the 16.2 fb^{-1} coming from the nominal dimuon trigger. The effect of prescaling is mitigated by including in the analysis events selected by two additional EF triggers scarcely affected by prescaling, where tighter selections were applied: $p_T > 6$ GeV or $|\eta| < 1.05$ for one of the muons. A full track reconstruction of the muon candidates was performed at the software trigger

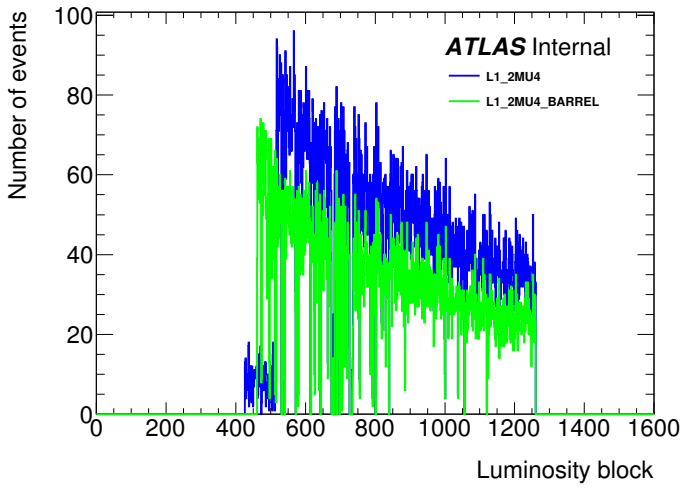

Figure 5.1: A plot of number of events recorded per luminosity block (a luminosity block refers to 1 minute of data-taking) for two triggers. The dimuon trigger used as the primary Level-1 trigger for this analysis (blue) has been prescaled for the first few hundred luminosity blocks of the data-taking run due to a large input rate of events. This is compared to an unprescaled L1 trigger (green), which had the additional requirement of the muons being in the ATLAS barrel ($|\eta| < 1.05$), leading to a lower event input rate and hence the trigger not requiring prescaling.

levels, where an additional loose selection was applied to the dimuon invariant mass $m_{\mu\mu}$ and the events were assigned to the J/ψ stream ($2.5 < m_{\mu\mu} < 4.3$ GeV) or to the B stream ($4.0 < m_{\mu\mu} < 8.5$ GeV). The J/ψ stream receives a much larger amount of data than the B stream, due to the lower mass range, and is thus prescaled by a factor of 10 - in order to keep synchronicity between the two streams, to be able to use the B^+ as a proxy for the B_s^0 in Section 5.7, the extra factor α_k in the master equation (5.1) is introduced as a normalisation term, to account for the different integrated luminosities.

Events from the 2012 dataset are divided into three mutually exclusive trigger categories:

- T_1: 'higher threshold' trigger with $p_T > 6$ GeV for one muon and > 4 GeV for the other one;
- T_2: 'barrel' trigger with $p_T > 4$ GeV for both muon candidates and at least one of them with $|\eta| < 1.05$ (and T_1 requirement not satisfied);
- T_3: basic dimuon trigger with $p_T > 4$ GeV for both muon candidates (and T_1, T_2 requirements not satisfied).

Events belonging to a given category are all associated with the same pattern of Level-1 prescaling. The T_1 category contains 81.3% of the data that passes one of the trigger requirements, the T_2 category 16.4%, and the T_3 category 2.3%. By combining the trigger categories in this exclusive manner, the

prescaling applied to the basic dimuon trigger at software level is mitigated, as shown in Figure 5.2. The impact of the Level-1 trigger prescaling on the total sample of collected events is minor, since the majority of the events belong to the T_1 category. The α_k factor can be calculated for each category from the luminosity information for each trigger; averaged across the full 2012 data-taking, these values are 7.23, 7.28 and 7.29 for T1, T2 and T3 respectively - this is mainly due to the prescaling of the J/ψ trigger stream. Figure 5.3 shows the distributions of several important variables in these categories, to show the differences between the selected events.

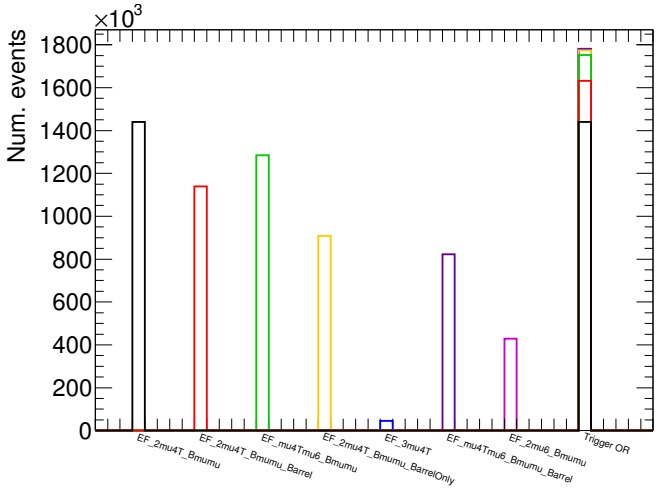

Figure 5.2: Number of recorded events for several di- or tri-muon triggers that were considered for use. The black trigger, EF_2mu4T_Bmumu, also known as the T3 category trigger, was prescaled at software level. By using the two other triggers, EF_2mu4T_Bmumu_barrel (T2, red) and EF_mu4Tmu6_Bmumu (T1, green), up to 98% of the luminosity lost by prescaling can be recovered.

A fourth category is defined for events from the 2011 dataset. They were collected with a trigger requirement $p_T > 4$ GeV for both muon candidates, and prescaling was not applied to this sample.

5.5 Background Composition

The background to $B_s^0 \to \mu^+\mu^-$ decays can be split into three types:

- *Combinatorial* background, consisting of pairs of uncorrelated muons produced from the same vertex, characterised by a small, approximately linear dependence on the dimuon invariant mass;

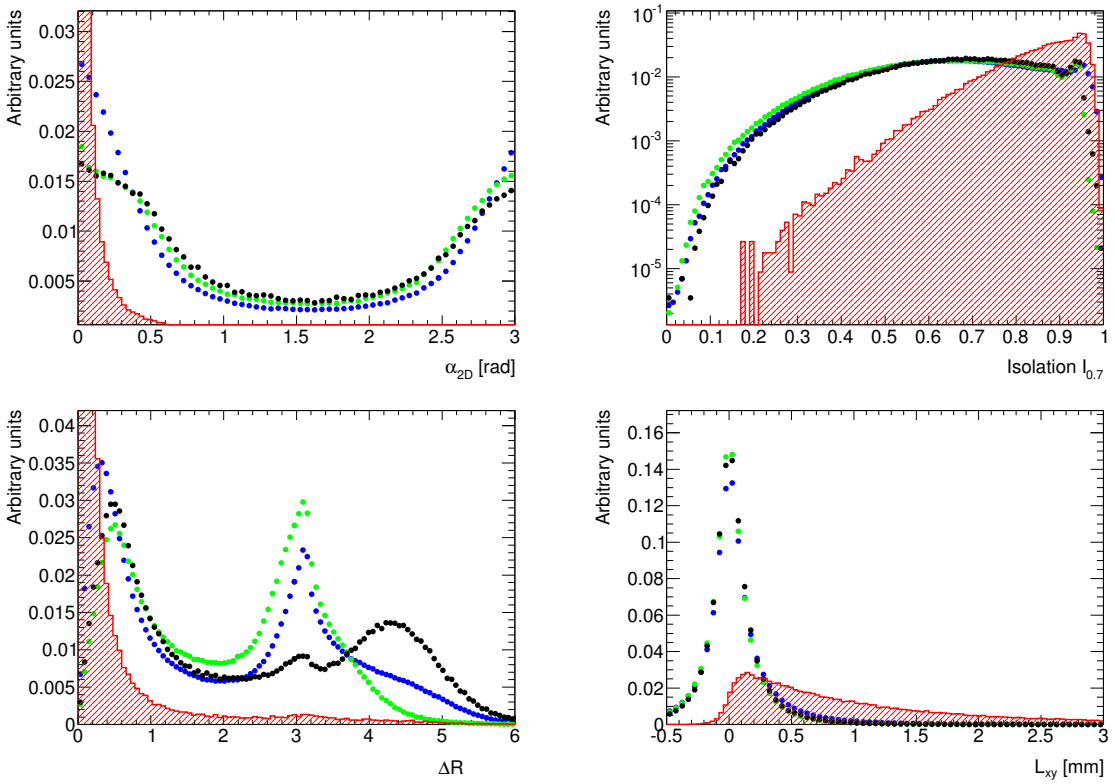

Figure 5.3: $B_s^0 \to \mu^+\mu^-$ sideband data (dots) separated into the three trigger categories - T1 (blue), T2 (green) and T3 (black). *From top to bottom, left to right:* distributions of α_{2D}, B meson isolation, ΔR and L_{xy}. The shape of the signal, taken from Monte Carlo, is shown in red for reference.

- *Partially reconstructed* $B \to \mu\mu X$ decays, where the decay of a B meson has not been reconstructed fully and the X component has been missed, resulting in two muons from the same decay vertex, generally sitting in the low-mass sideband;
- *Peaking* backgrounds, arising from $B^0_{(s)} \to hh'$ decays, where both hadrons are misindentified as muons. This is dangerous as the (apparent) dimuon invariant mass distribution of these events peaks within the signal region, due to them originating from a real B_s^0, and cannot be observed in the sidebands, making it virtually indistinguishable from the signal.

The combinatorial background is mostly caused by muons produced independently by the decay and fragmentation of a b and a $\bar{b}$ quark, also known as *opposite-side* muons. It is studied in the mass sidebands in the $B_s^0 \to \mu^+\mu^-$ channel, and Section 5.7 contains comparisons between data and MC that show that this background is mostly well-modelled and understood; some discrepancies between

the two do exist, and are due to residual differences between the MC and the data. The multivariate classifier discussed in Section 5.6 is used primarily to reduce this background.

Partially reconstructed decays can themselves be sorted into three types:

(a) *Same-side (SS)* combinatorial background where a muon pair originates from the same b-quark, e.g. $b \to c\mu\nu \to s(d)\mu\mu\nu\nu$, where only the muon pair is visible to the detector;

(b) *Same-vertex (SV)* background from B decays with a muon pair coming from the same vertex, e.g. $B^0 \to K^{*0}\mu\mu$, or $B \to J/\psi X \to \mu\mu\mu X'$;

(c) B_c decays such as $B_c \to J/\psi\mu\nu \to \mu\mu\mu\nu$, where the third muon is either missed or misreconstructed;

(d) *Semileptonic* b-hadron decays, where the final state hadron is misidentified as a muon.

Inclusive MC samples for the SS/SV backgrounds and the $B_c \to J/\psi\mu\nu$ decays were produced together with a very large combinatorial background sample. The dimuon mass range for the partially-reconstructed decay samples is mostly below the $B_s^0 \to \mu^+\mu^-$ mass window, however the high mass tail does extend into the signal region and becomes a significant contributor to the total background in the final selection, after the continuum-reducing classifier discussed in Section 5.6 is applied. The semileptonic component consists mainly of 3-body charmless decays, such as $B_s^0 \to K\mu\nu$ or $\Lambda_b \to p\mu\nu$, where the hadron is misidentified as a muon, and the tail of the mass distribution extends into the signal region. These kinds of decays do not provide a large contribution to the total background, due to branching fractions of the order of 10^{-6}, and their contribution is further reduced by the dedicated muon identification requirements discussed in Section 5.3.

5.6 Multivariate Classifiers for Reducing Background

This analysis utilises two multivariate classifiers to reduce backgrounds. The first is used to discriminate against events where the muon pair did not originate from a secondary vertex. The second is used to drastically reduce the dominant continuum background. While the continuum-reducing classifier was used in the previous analysis on the 2011 data [65], the fake muon classifier is a new addition, and the continuum classifier is also improved with larger MC samples compared to 2011.

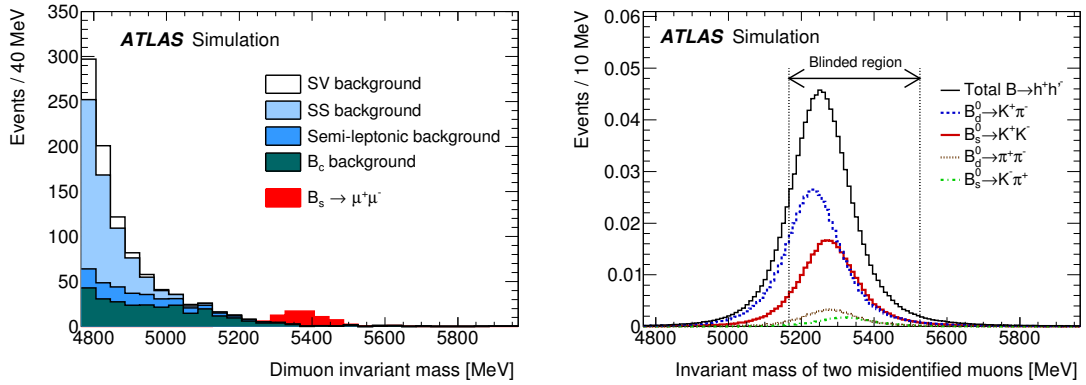

Figure 5.4: (a) Dimuon mass distribution for the partially reconstructed decays, taken from MC, after the candidate selection detailed in Section 5.3 is applied. The four components of these decays are shown as stacked histograms, and all are normalised according to the world-average measured branching fractions. The SM expectation for the $B_s^0 \to \mu^+\mu^-$ signal is shown in red for reference. (b) Dimuon invariant mass distributions for the various contributions to the peaking background, after candidate selection and both the fake-reducing multivariate classifier and the continuum-reducing classifier detailed in Section 5.6 are applied. Both plots are normalised to expected yields for a total integrated luminosity of 25 fb^{-1}. [58]

Multivariate Analysis for Fake Muons

Muons are among the easiest objects to reconstruct with the ATLAS detector - tracks appear in both the ID and MS systems, in line with each other, with no energy deposits in the calorimeter. The performance of the ATLAS muon reconstruction is reported in [63]. Thanks to a full `Geant-4` [62] simulation of the ATLAS detector, a study was performed to find the rate of particle misidentification. The majority of these so-called 'fake' muons actually come from real muons, originating from decays in flight of kaons and pions, where the muon receives most of the energy of the meson. These real muons are denoted as fake only by nature of not coming directly from a secondary vertex - for example, the J/ψ from a $B_s^0 \to J/\psi\,\phi$ decay produces two muons, which are labelled as fake only because they did not come directly from the B_s^0 decay vertex. Hadronic 'punch-through', where a hadron does not lose all its energy in the calorimeter and makes it through into the MS, contributes just 3% (8%) to the fake rate from kaons (pions). After applying the preliminary selection, this contribution drops to just 0.4% (0.2%) for kaons (pions). Protons have a negligible contribution ($< 0.01\%$).

A Boosted-Decision Tree (BDT) is used as the multivariate classifier for this study, as it was found to be the best performing. The BDT training was done using the multivariate analysis tool TMVA [66], using the simulated $B_s^0 \to \mu^+\mu^-$ and $B \to hh'$ samples mentioned previously. The variables used in this BDT are described in Table 5.1, and are quantities related to the ID and MS and the quality of

the reconstruction performed for both systems.

1.	Absolute value of the track rapidity measured in the ID.
2.	Ratio q/p (charge over momentum) measured in the MS.
3.	Scattering curvature significance: maximum value of the significance of the track curvature variation across each layer of the ID.
4.	χ^2 of the track reconstruction in the MS.
5.	Number of hits used to reconstruct the track in the MS.
6.	Ratio of the values of q/p measured in the ID and in the MS, corrected for the average energy loss in the calorimeter.
7.	χ^2 of the match between the tracks reconstructed in the ID and MS.
8.	Energy deposited in the calorimeters along the muon trajectory obtained by combining ID and MS tracks.

Table 5.1: Description of the eight variables used in the discrimination between signal muons and those from hadron decays in flight and punch-throughs.

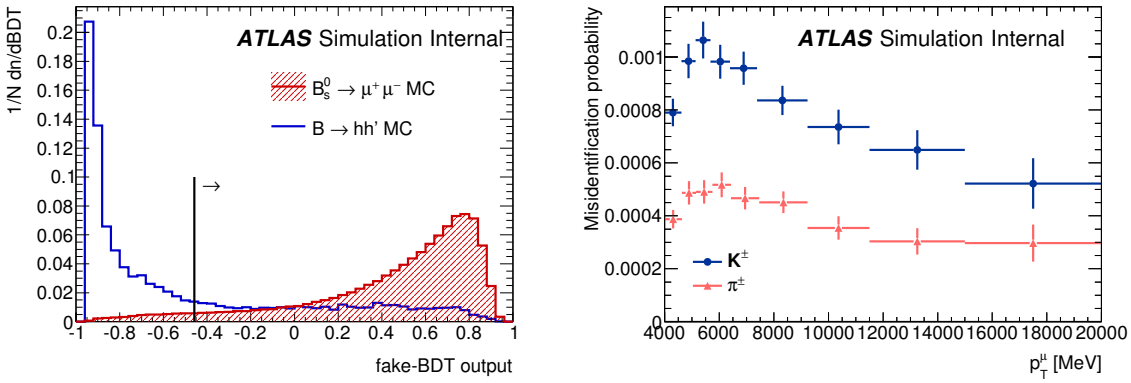

Figure 5.5: *Left*: BDT output for the fake-BDT, with the real muons from the $B_s^0 \to \mu^+\mu^-$ signal shown in red and fake muons from the peaking $B \to hh'$ background in blue. *Right*: Misidentification probabilities for the charged kaons and pions, as a function of their p_T.

The fake-BDT selection chosen, optimised based on the $B_s^0 \to \mu^+\mu^-$ significance and shown in Figure 5.5, retains a 95% efficiency for the real muons in the signal sample, and achieves an average reduction of hadron misidentification by a factor of 2.7. This results in a final misidentification probability of 0.09% for kaons, and 0.04% for pions, also shown in Figure 5.5.

The peaking background due to $B_s^0 \to hh'$ events with double hadron misidentification is extremely hard to distinguish from the $B_s^0 \to \mu^+\mu^-$ signal, given that it sits in the same mass window. These

events have the same acceptance and selection efficiency as the signal, and as the misidentified hadrons both originate from the secondary vertex, the fake-BDT classifier does not help reduce this background. However, in much the same way as the $B_s^0 \to \mu^+\mu^-$ signal itself, this background contribution can be estimated from the number of observed $B^+ \to J/\psi\, K^+$ events using Eq. 5.1. Using the world average branching fraction values for B^0 and B_s^0 decays into KK, $K\pi$ and $\pi\pi$ [3], along with the hadron misidentification probabilities previously calculated, the expected number of these peaking background events after final selection (including the continuum-BDT discussed in the next section) is 0.7, with a 10% uncertainty from the normalisation procedure. The simulated samples of $B^+ \to J/\psi\, K^+$ were validated and calibrated by comparing to data; a factor 1.2 ± 0.2 difference between the two was found, which is used a scale correction and systematic uncertainty, leading to a final expected peaking-background estimation of 1.0 ± 0.4 events.

Reducing the peaking background is not the only effect of this fake-BDT selection - the semileptonic background, arising from decays such as $B^0 \to \pi\mu\nu$ and $B_s^0 \to K\mu\nu$ with a misidentified hadron, is also reduced by this selection. Taken from MC, the expected number of these events in the final selection is 107 ± 27. The equivalent $\Lambda_b \to p\mu\nu$ contribution is negligible, due to a smaller production cross section, and the fact that ATLAS has a fake rejection rate for proton identification at the level of 10^{-5} [58]. This compares favourably to the CMS and LHCb result, where proton misidentification occurs at approximately the same rate as kaon and pion misidentification, $\mathcal{O}(10^{-3})$, and is the leading source of branching fraction systematic uncertainty [57].

Multivariate Classifier for Continuum Reduction

The dominant source of background in this analysis is the continuum background, described in Section 5.5. With just the baseline selection detailed in Section 5.3 and the fake-BDT selection described above, the continuum would dominate the signal with approximately 1 signal event per 10^6 continuum events, making detection and measurement of $B_s^0 \to \mu^+\mu^-$ decays nearly impossible. Hence in both the previous analysis [65] and this analysis [58], a BDT is used to dramatically reduce this background. This classifier was trained on 15 variables, described in Table 5.2, which can be split into three groups: (a) B meson variables related to the reconstruction of the decay vertex and correlations between $\vec{p}^B$ and the production and decay vertices; (b) variables describing the muons coming from the B meson candidate; and (c) variables related to the rest of the event surrounding the B meson candidate. These variables were chosen to optimise the discriminating power of the classifier, while minimising the dependence on the muon pair invariant mass.

The correlations between variables in the signal sample were studied, as well as in the continuum background on both MC and the sideband data. The variables $\chi^2_{PV,DVxy}$, L_{xy}, $|d_0|^{max}$ and $|d_0|^{min}$

Variable	Description
p_T^B	Magnitude of the B candidate transverse momentum $\overrightarrow{p_T}^B$.
$\chi^2_{\mathrm{PV,DV}\ xy}$	Significance of the separation $\overrightarrow{\Delta x}$ between production (*i.e.* associated PV) and decay (DV) vertices in the transverse projection: $\overrightarrow{\Delta x}_{\mathrm{T}} \cdot \Sigma^{-1}_{\overrightarrow{\Delta x}_{\mathrm{T}}} \cdot \overrightarrow{\Delta x}_{\mathrm{T}}$, where $\Sigma_{\overrightarrow{\Delta x}_{\mathrm{T}}}$ is the covariance matrix.
ΔR	three-dimensional opening between $\overrightarrow{p}^B$ and $\overrightarrow{\Delta x}$: $\sqrt{{\alpha_{\mathrm{2D}}}^2 + \Delta\eta^2}$.
$\lvert\alpha_{\mathrm{2D}}\rvert$	Absolute value of the angle between $\overrightarrow{p_T}^B$ and $\overrightarrow{\Delta x}_{\mathrm{T}}$ (transverse projection).
L_{xy}	Projection of $\overrightarrow{\Delta x}_{\mathrm{T}}$ along the direction of $\overrightarrow{p}^B_{\mathrm{T}}$: $(\overrightarrow{\Delta x}_{\mathrm{T}} \cdot \overrightarrow{p_T}^B)/\lvert\overrightarrow{p_T}^B\rvert$.
$\mathrm{IP}_B^{\mathrm{3D}}$	three-dimensional impact parameter of the B candidate to the associated PV.
$\mathrm{DOCA}_{\mu\mu}$	Distance of closest approach (DOCA) of the two tracks forming the B candidate (three-dimensional).
$\Delta\phi_{\mu\mu}$	Difference in azimuthal angle between the momenta of the two tracks forming the B candidate.
$\lvert d_0\rvert^{\mathrm{max}}$-sig.	Significance of the larger absolute value of the impact parameters to the PV of the tracks forming the B candidate, in the transverse plane.
$\lvert d_0\rvert^{\mathrm{min}}$-sig.	Significance of the smaller absolute value of the impact parameters to the PV of the tracks forming the B candidate, in the transverse plane.
$P_{\mathrm{L}}^{\mathrm{min}}$	Value of the smaller projection of the momenta of the muon candidates along $\overrightarrow{p_T}^B$.
$I_{0.7}$	Isolation variable defined as ratio of $\lvert\overrightarrow{p_T}^B\rvert$ to the sum of $\lvert\overrightarrow{p_T}^B\rvert$ and of the transverse momenta of all additional tracks contained within a cone of size $\Delta R < 0.7$ around the B direction. Only tracks with $p_T > 0.5$ GeV and matched to the same PV as the B candidate are included in the sum.
$\mathrm{DOCA}_{\mathrm{xtrk}}$	DOCA of the closest additional track to the decay vertex of the B candidate. Tracks matched to a PV different from the B candidate are excluded.
$N_{\mathrm{xtrk}}^{\mathrm{close}}$	Number of additional tracks compatible with the decay vertex (DV) of the B candidate with $\ln(\chi^2_{\mathrm{xtrk,DV}}) < 1$. The tracks matched to a PV different from the B candidate are excluded.
$\chi^2_{\mu,\mathrm{xPV}}$	Minimum χ^2 for the compatibility of a muon in the B candidate with a PV different from the one associated with the B candidate.

Table 5.2: Description of the 15 variables used in the discrimination between signal and continuum background. When the BDT classifier is applied to $B^+ \to J/\psi\, K^+$ and $B_s^0 \to J/\psi\, \phi$ candidates, the variables related to the decay products of the B mesons refer only to the muons from the decay of the J/ψ. Variables appear in order of their importance according to the classifier training.

significance and $\chi^2_{\mu,xPV}$ have significant linear correlations with each other - unsurprising perhaps, as they all relate to the difference between primary and decay vertices. On the other hand, the variables IP_B^{3D}, $\mathrm{DOCA}_{\mu\mu}$ and $I_{0.7}$ have negligible correlations with any other variables, thus providing unique information and discrimination between signal and background.

BDT training and testing was done using TMVA [66] using the $B_s^0 \to \mu^+\mu^-$ signal MC sample and the very large MC sample of $B^0 \to hh'$ semileptonic hadron decays. The resulting output is shown in Figure 5.6, for the signal and different background samples, with 2012 sideband data, taken from the low- and high-mass regions in the B_s dimuon invariant mass distribution, shown overlaid for reference. For both signal and background, the linear correlation between the BDT output distribution and the dimuon invariant mass is less than 1%. The chosen BDT selection requires the output value to be greater than 0.24, resulting in a 54% relative signal efficiency and a reduction in the combinatorial background by a factor $\mathcal{O}(1000)$.

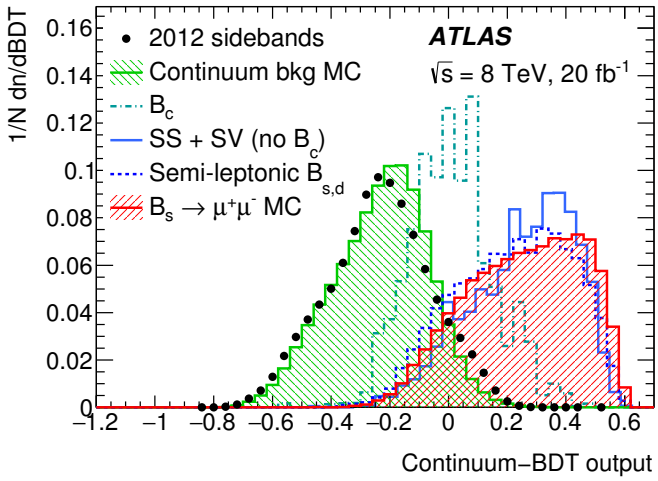

Figure 5.6: Continuum-BDT output distribution for signal and background events, with sideband data overlaid for reference. The $B_s^0 \to \mu^+\mu^-$ signal is shown as the solid red histogram, and the combinatorial background as the solid green histogram. All distributions are shown with only the baseline selection from Section 5.3 applied, and before applying any reweighting to the discriminating variables used in the classifier training.

5.7 Data-Monte Carlo Comparisons

Comparisons between data and MC are used to assess the data-driven weighting of the MC samples, and to evaluate any systematic differences between the two. This has been done using the discriminating

variables from the continuum-BDT, with the data from the $B_s^0 \to \mu^+\mu^-$ sidebands and the $B^+ \to J/\psi\, K^+$ and $B_s^0 \to J/\psi\, \phi$ channels being compared to the equivalent MC samples. In all cases, the baseline selection detailed in Section 5.3 has been applied along with the fake-BDT selection, but the continuum-BDT selection has not been required.

$B_s^0 \to \mu^+\mu^-$ Continuum

The large MC sample of combinatorial background is compared to the sideband data for several discriminating variables. The sideband data is expected to contain mostly combinatorial events, with a small contribution from partially-reconstructed decays. Figure 5.7 shows these distributions - agreement between the sideband data and MC is fair, and though there are some discrepancies (the largest by far of which is in the $\chi^2_{\mu,xPV}$ variable shown in Figure 5.7), the continuum MC is used only for the training of the continuum-BDT classifier, and the differences can at worst lead to a non-optimal classifier.

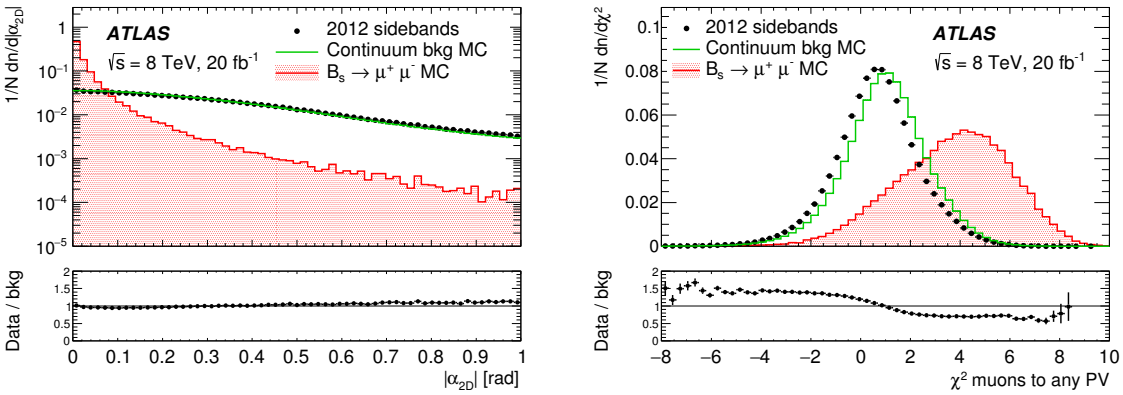

Figure 5.7: Data and MC continuum distributions for $|\alpha_{2D}|$ (*left*) and $\chi^2_{\mu,xPV}$ (*right*) variables (see Table 5.2). The 2012 sideband data is shown as the black dots, with the continuum MC distribution shown as the green histogram. The $B_s^0 \to \mu^+\mu^-$ signal MC is provided in red as reference.

$B^+ \to J/\psi\, K^+$ and $B_s^0 \to J/\psi\, \phi$ Reference Channels

The same comparison must be carried out in the reference channels. MC samples for both $B^+ \to J/\psi\, K^+$ and $B_s^0 \to J/\psi\, \phi$ are compared to data - however, as the data contains both the signal and the background in each channel, the background must be subtracted before the comparison can be done. This background subtraction is performed using a binned maximum-likelihood fit to the dimuon invariant mass distribution, with one fit per trigger category.

For the B^+ the signal is described by a single Gaussian distribution in the T2 and T3 trigger categories and by a double-Gaussian distribution in the T1 and 2011 categories. An error function is used to fit the partially-reconstructed decays, and an exponential function is used for the continuum background. For the $B_s^0 \to J/\psi\,\phi$ channel, a Gaussian distribution is used for the signal with a third-order Chebychev polynomial for the background. A few extra selection cuts are necessary to address the additional kaon in the $J/\psi\phi$ final state: both kaons must have $p_T > 1$ GeV, and to select the ϕ mass window, the invariant mass of the two hadrons must be within 15 MeV of the ϕ mass. Figure 5.8 shows the fitted mass distributions in the B^+ channel for each trigger category.

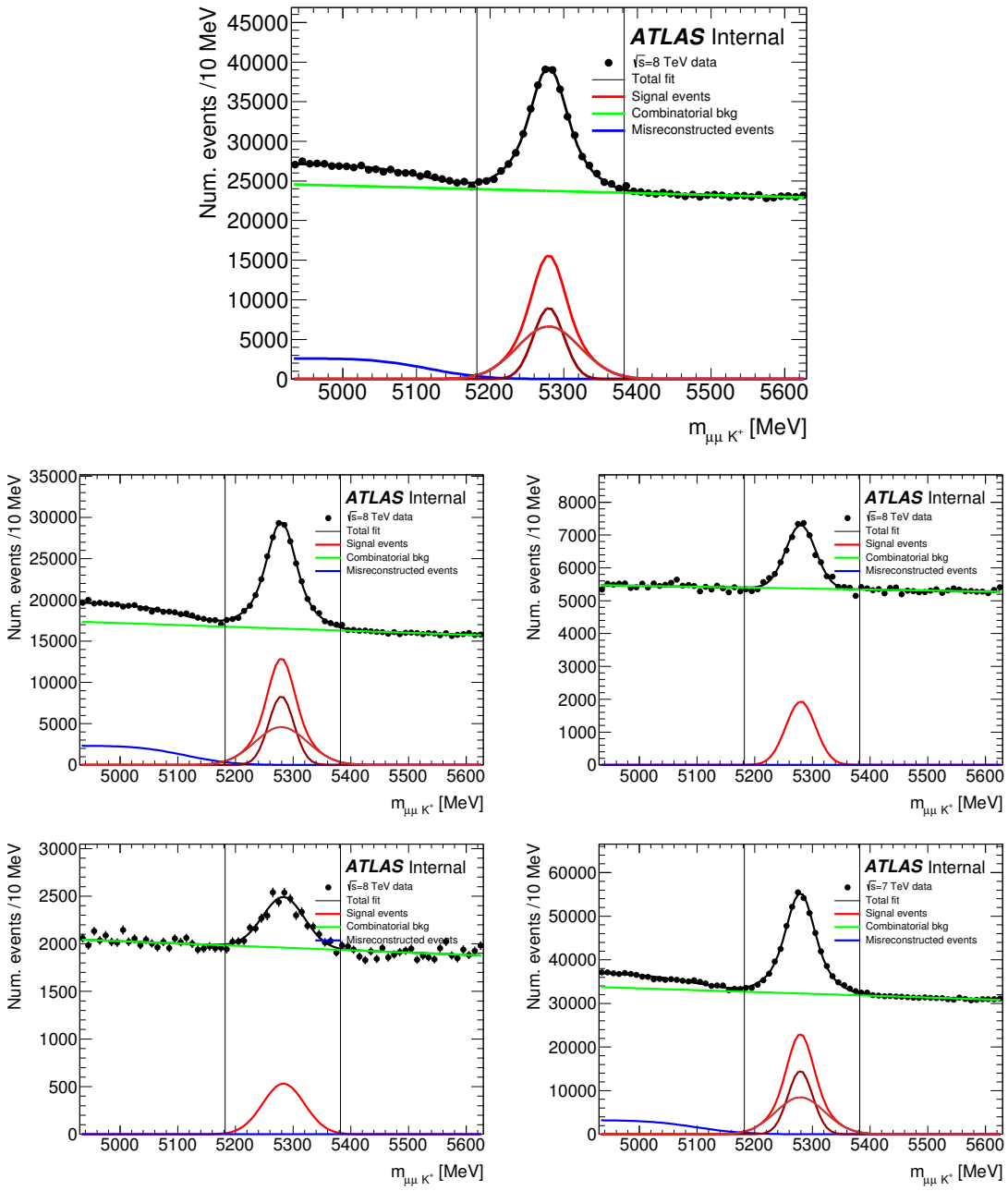

Figure 5.8: $B^+ \to J/\psi\, K^+$ dimuon invariant mass distributions for (*left to right, top to bottom*): all 2012 data combined, T1, T2, T3 and 2011 trigger categories. The background subtraction is done using the combined 2012 data, with the shape of the various background contributions taken from the fit shown in the top plot.

To do the subtraction, the sideband data in each channel ($4930 < m_{J/\psi K^+} < 5180$ and $5380 < m_{J/\psi K^+} < 5630$ for the B^+, and $5050 < m_{J/\psi\phi} < 5297$ and $5437 < m_{J/\psi\phi} < 5650$ for the B_s) is used to estimate the number of background events in the signal mass window, and the background shape in each variable. In the case of the B^+, this is made a little more complicated, as the partially reconstructed decays accumulate in the low-mass sideband, causing an over-estimation of the continuum in the signal region. As a result, only the combinatorial background is considered in the low-mass sideband, which is reweighted accordingly. The partially-reconstructed decays in that region were shown, using MC samples, to have the same shape as the $B^+ \to J/\psi\, K^+$ signal in the discriminating variables considered, so there is no distortion of signal shape caused by not subtracting these background events.

Figures 5.9, 5.10 and 5.11 show some of the discriminating variable distributions for both reference channels after the background subtraction. In general the agreement is good - the largest discrepancy is in the B^+ channel, in the B isolation variable $I_{0.7}$. The differences between data and MC in the B^+ channel are accounted for as systematics during the evaluation of the $B_s^0 \to \mu^+\mu^-$ vs $B^+ \to J/\psi\, K^+$ efficiency×acceptance ratio, detailed in Section 5.9. Figure 5.12 shows the continuum-BDT output value for both channels - the agreement for the $B_s^0 \to J/\psi\,\phi$ channel is excellent, however there is a large discrepancy in the B^+ channel. This discrepancy is assessed during the ratio evaluation in Section 5.9.

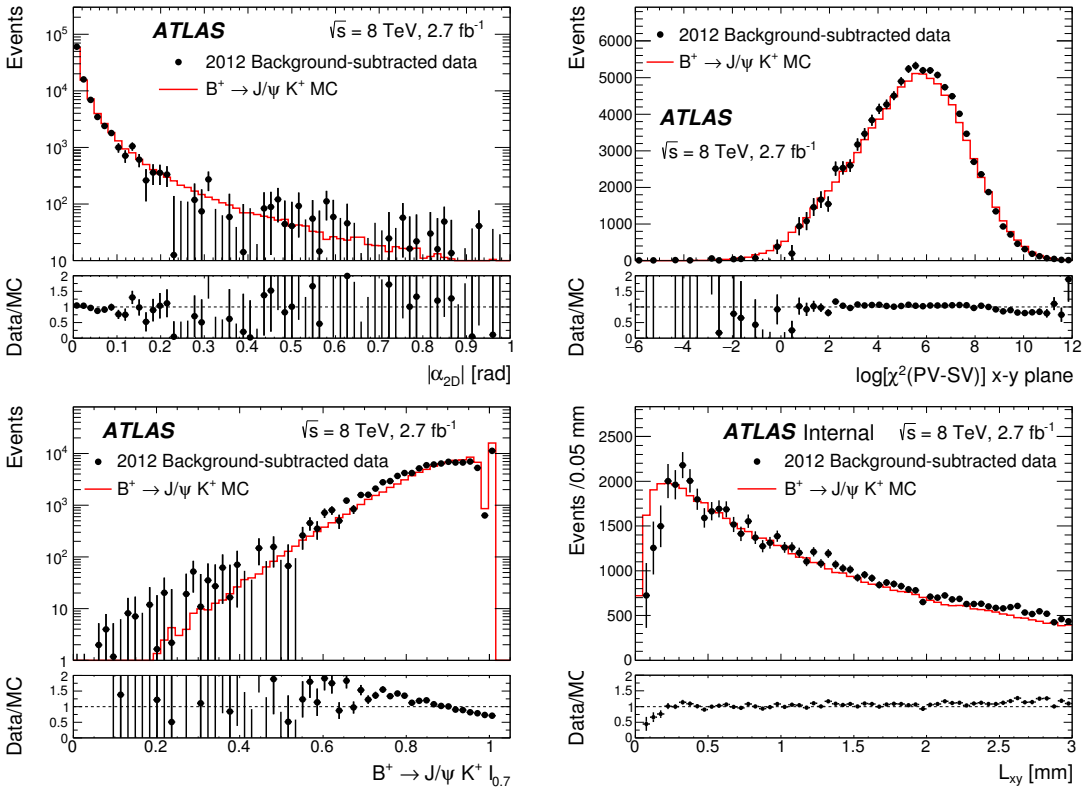

Figure 5.9: Distributions of some of the discriminating variables in the $B^+ \to J/\psi\, K^+$ reference channel. The data has been background-subtracted, leaving the $B^+ \to J/\psi\, K^+$ signal and signal-like partially-reconstructed decays. The MC signal sample is shown in red. No continuum-BDT cut has been applied. *From left to right, top to bottom*, the discriminating variables are: α_{2D}, $\chi^2_{\mathrm{PV,DV}\,xy}$, B meson isolation, and the transverse decay length L_{xy}.

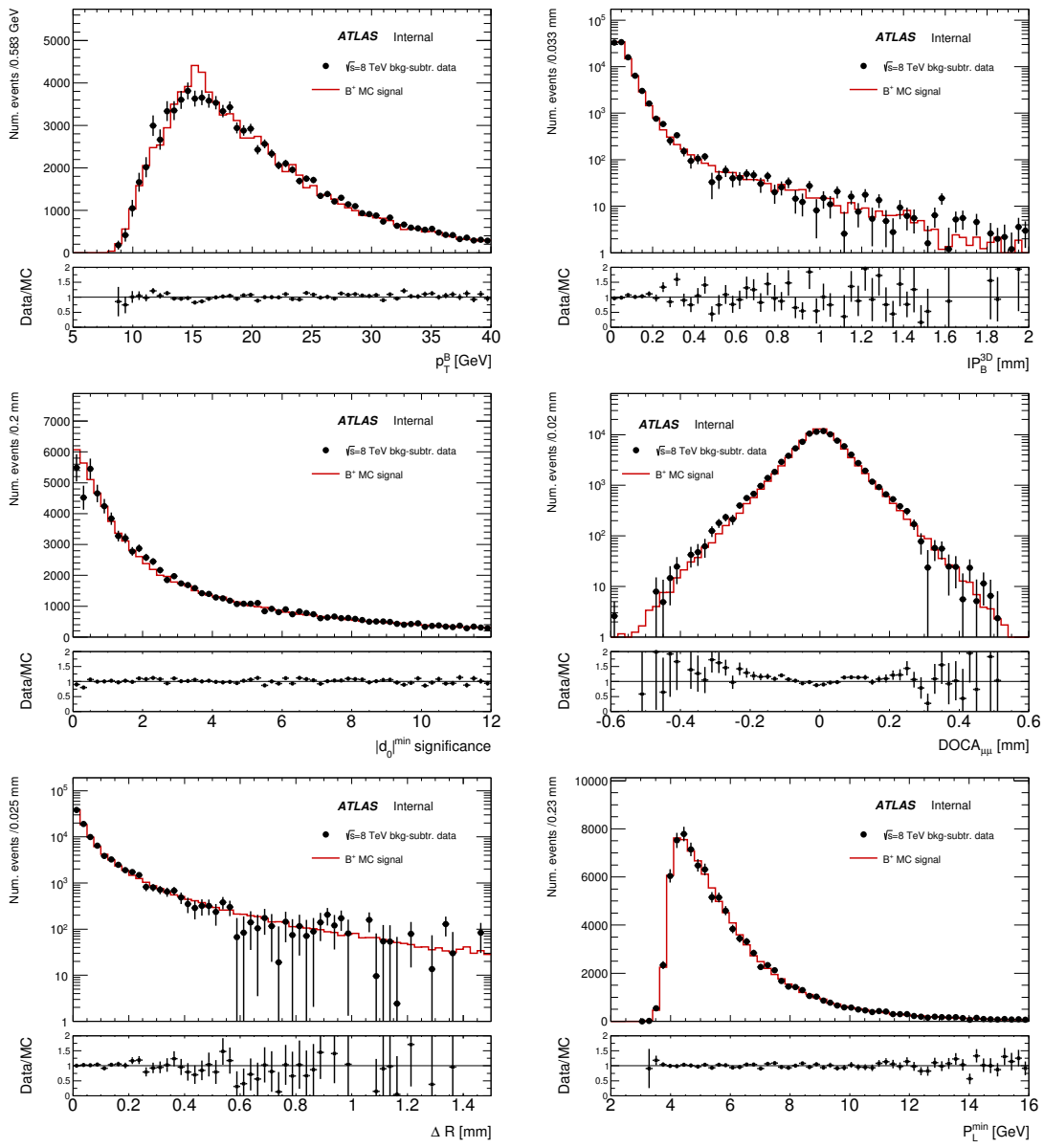

Figure 5.10: Distributions of some of the discriminating variables in the $B^+ \to J/\psi\, K^+$ reference channel. The data has been background-subtracted, leaving the $B^+ \to J/\psi\, K^+$ signal and signal-like partially-reconstructed decays. The MC signal sample is shown in red. No continuum-BDT cut has been applied. *From left to right, top to bottom*, the discriminating variables are: p_T^B, IP_B^{3D}, $|d_0|^{\min}$ significance, $\Delta\, R$ and $P_L^{\min}$.

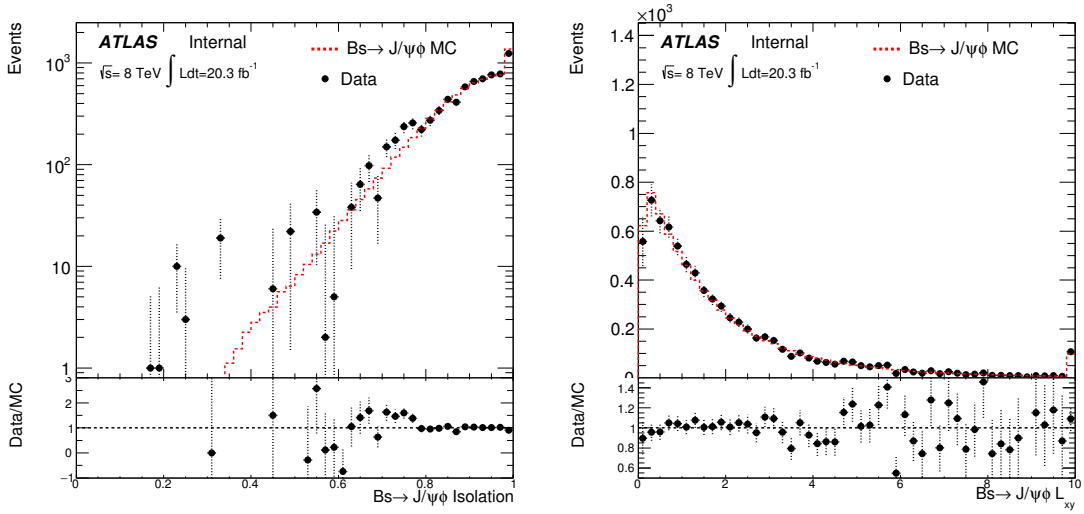

Figure 5.11: Distributions of B meson isolation (*left*) and the transverse decay length L_{xy} (*right*) for the $B_s^0 \to J/\psi\,\phi$ reference channel. Background-subtracted data is shown as the black dots, and the signal MC as the red histogram. No continuum-BDT cut has been applied.

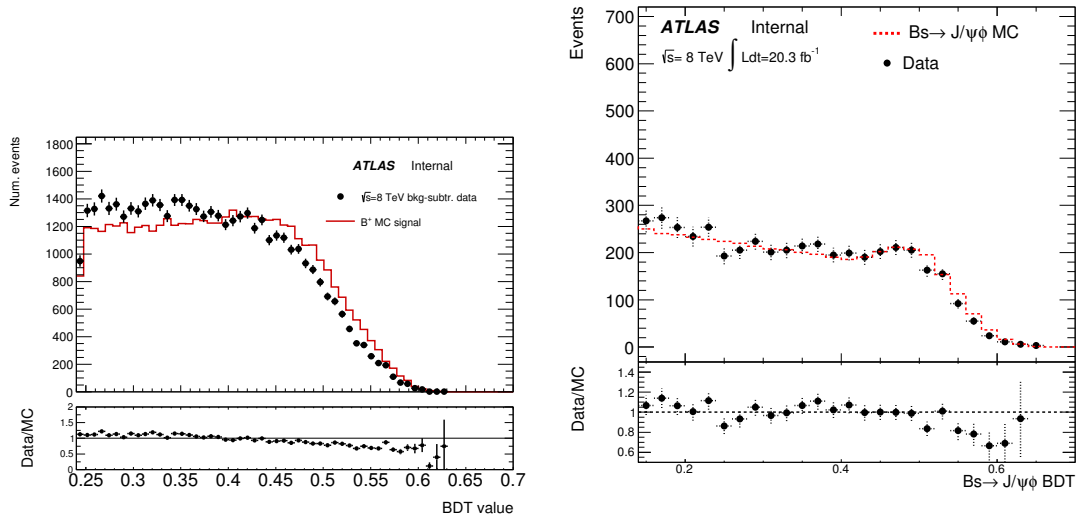

Figure 5.12: BDT output distributions, in the selected BDT region. *Left*: $B^+ \to J/\psi\,K^+$ distribution - the discrepancy here is dealt with by reweighting the distribution according to a linear function, details of which are found in Section 5.9. *Right*: $B_s^0 \to J/\psi\,\phi$ channel, where agreement is good and no reweighting is required.

5.8 $B^+ \to J/\psi\, K^+$ Yield Extraction

While the binned fit done in the previous section is good enough for a data-MC comparison, a more precise description of the background is required to accurately extract the $B^+ \to J/\psi\, K^+$ yield. Instead an unbinned extended maximum-likelihood fit is done, again on the $J/\psi K^+$ mass distribution. This fit includes four components:

- $B^+ \to J/\psi\, K^+$ events, parameterised by the combination of a Johnson S_U function (a transformation of a normal distribution with asymmetric tails) [67, 68] and a Gaussian function in the T1, T2 and 2011 trigger categories and by a single Johnson S_U function for the T3 category; this was modelled by a double (single) Gaussian function in the T1 and 2011 (T2 and T3) categories in the binned fit;
- Cabibbo-suppressed $B^+ \to J/\psi\, \pi$ events on the high-mass tail of the peak, modelled by the sum of a Johnson S_U and a Gaussian function - not included in the binned fit;
- Partially-reconstructed decays (PRD), described with a combination of Fermi-Dirac and exponential functions, compared to a simple error function in the binned fit;
- and the combinatorial background of mostly $b\bar{b} \to J/\psi\, X$ events, modelled by an exponential function, the only component described in the same way in both binned and unbinned fits.

The functions used to parameterise these contributions were obtained from studies of MC samples. All the yields are extracted from the fit to data, while the shape parameters are driven by the MC directly. The mass scale and mass resolution are added as free parameters, to allow for data-MC differences. As an example, the fit to the T1 category is shown in Figure 5.13. The results for all four trigger categories are shown in Table 5.3.

Category	$N_{J/\psi K^+}$ [events]	$N_{J/\psi \pi^+}$ [events]
T_1	$46\,860 \pm 290 \pm 280$	$1\,420 \pm 230 \pm 440$
T_2	$5\,200 \pm 84 \pm 100$	$180 \pm 51 \pm 89$
T_3	$2\,512 \pm 91 \pm 42$	$85 \pm 77 \pm 30$
2011	$95\,900 \pm 420 \pm 1\,100$	$3\,000 \pm 340 \pm 1\,140$

Table 5.3: Results of the fits to the events reconstructed as $B^+ \to J/\psi\, K^+$ in each trigger category. Uncertainties are statistical and systematic, respectively. [58]

The fit itself accounts for some systematic effects, such as the limited MC sample size, which is included through the simultaneous fit to data and MC. Differences in mass scale and resolution, the

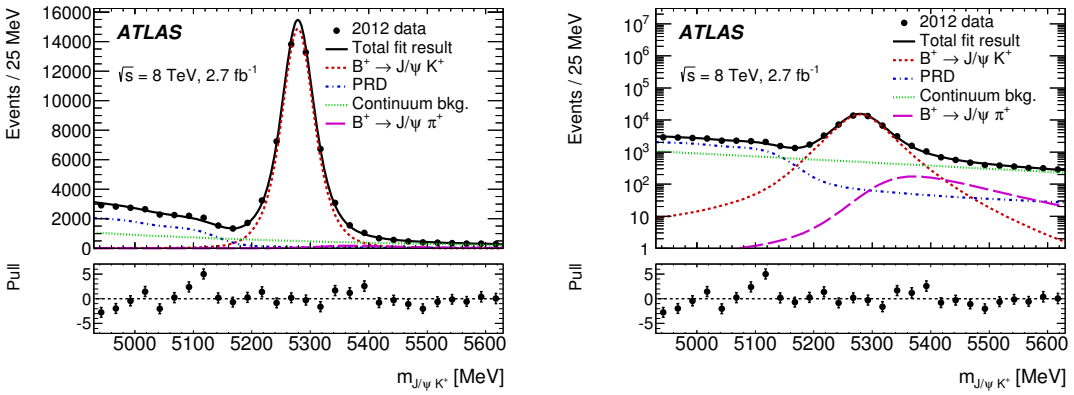

Figure 5.13: $J/\psi\, K^+$ invariant mass distribution for all B^+ candidates in the T1 trigger category, in linear (*left*) and logarithmic (*right*) scale. The subplot underneath shows the bin-by-bin pulls for the fit, where the pull is defined as the difference between the data point and the value obtained from the fit function, divided by the error from the fit. [58]

free parameters in the fit, between data and MC are accounted for by obtaining scaling factors from the fit. Other systematic uncertainties, such as differences in efficiency between opposite-charged B decays, uncertainties in the relative fractions and shapes of the partially-reconstructed decays, and uncertainties in the shape of the combinatorial background, are evaluated by varying the default fit model. The difference between the default and varied fits are symmetrised and used as an estimate of the systematic uncertainty. In total the statistical and systematic uncertainty in the B^+ normalisation yield is 0.8%.

5.9 Evaluation of the $B_s^0 \to \mu^+\mu^-$ to $B^+ \to J/\psi\, K^+$ Efficiency Ratio

The branching ratio for $B_s^0 \to \mu^+\mu^-$ decays is not measured directly - as explained in the introduction to this chapter, it is measured relative to the known branching ratio for $B^+ \to J/\psi\, K^+$. The ratio of the efficiencies enters into the branching fraction equation in the $\mathcal{D}_{norm}$ term, as detailed in Eq. 5.2. Both channels are measured in the same fiducial volume, defined in terms of the B meson properties $p_T^B > 8.0$ GeV and $|\eta_B| < 2.5$. The efficiencies defined with respect to this fiducial volume include both the acceptance and the selection efficiency; the acceptance is defined as the ratio between the number of events passing the fiducial volume selection and the number of events passing the final state particle selection, which requires $p_T^\mu > 4$ GeV and $|\eta_\mu| < 2.5$ for muons and $p_T^K > 1$ GeV and $|\eta_K| < 2.5$ for kaons. The acceptance is evaluated using a specific 'un-biased' MC sample where only the fiducial volume cuts on the B meson are applied.

The selection efficiency covers the trigger, reconstruction and signal selection efficiencies. Trigger efficiencies were taken from a data-driven study [69], based on comparing single- and di-muon triggers for events containing muon pairs from J/ψ and Υ resonance decays. The signal selection efficiency is defined as the ratio between the number of events passing the final state particle selection (detailed above) and the number of events passing all cuts up to and including the continuum-BDT > 0.24 requirement. All efficiency terms are computed separately in the three 2012 trigger selections and the 2011 selection, using the nominal MC samples.

Table 5.4 shows the efficiency ratios R_ϵ^k for each trigger category, labelled $k = 1-4$, with the statistical and systematic uncertainties as well as the relative contribution to the normalisation term $\mathcal{D}_{norm}$. The same ratios are used in both $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching ratio calculations. The systematic uncertainties contributing to these values are summarised in Table 5.5.

Data category (k)	$R_\epsilon^k = (\epsilon_{J/\psi K^+}/\varepsilon_{\mu^+\mu^-})_k$	Relative contribution to $\mathcal{D}_{\text{norm}}$
T_1	$0.180 \pm 0.001 \pm 0.009$	68.3%
T_2	$0.226 \pm 0.004 \pm 0.014$	6.0%
T_3	$0.189 \pm 0.005 \pm 0.022$	3.5%
2011	$0.156 \pm 0.002 \pm 0.009$	22.2%

Table 5.4: Values of the efficiency ratios R_ϵ^k for the 2012 trigger categories and the 2011 sample, and their relative contributions to $\mathcal{D}_{\text{norm}}$ (Eq. (5.2)). The first uncertainty is statistical and the second systematic. The systematic component includes the uncertainties from the MC reweighting and from data–MC discrepancies, as described in the text.

Statistical uncertainty in simulation	0.5%
p_T, η reweighting	0.8%
Trigger efficiency	1.9%
Data to MC discrepancy in discriminating variables	4.2%
K^+ and B^+ reconstruction	3.6%
B^+ yield	0.8%
Total uncertainty	5.9%

Table 5.5: Summary of the systematic uncertainties in the $\mathcal{D}_{\text{norm}}$ term of Eq. (5.2).

While the statistical uncertainty arises purely because of the limited number of events in the simulated samples, the systematic uncertainty on R_ϵ^k comes from four sources:

(1) the uncertainties due to corrections applied to the MC - namely, the DDW and the trigger efficiencies;

(2) data-MC differences in the distributions of the discriminating variables, as evaluated in Section 5.7;

(3) differences between the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\, K^+$ channels related to the reconstruction efficiency of the kaon track and of the B^+ decay vertex;

(4) residual uncertainties in the trigger efficiencies.

The first of these, the uncertainties due to MC reweighting and trigger efficiencies, is assessed by pseudo-MC studies, varying the corrections within their statistical uncertainties. The RMS value of the distribution of the resulting R_ϵ^k values is taken as a systematic uncertainty; this ranges from $\pm 1\%$ to $\pm 5\%$ depending on the trigger category, with a combined effect of $\pm 1.5\%$ for the 2012 data and $\pm 2.2\%$ for 2011.

The second source of systematic uncertainty is evaluated by reweighting the MC samples for both $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\, K^+$ channels in each discriminating variable, according to the distribution of B^+ events observed in data after the background subtraction described in Section 5.7. The difference between the efficiency ratio values before and after this reweighting is taken as the systematic uncertainty due to mis-modelling of that variable. The single largest contribution to this uncertainty comes from the isolation variable $I_{0.7}$ (see Table 5.2 for the definition), which is computed using charged-particle tracks only and hence differences between B^+ and B_s^0 are expected. For only this variable, $B_s^0 \to J/\psi\, \phi$ decays are used as a substitute for the signal, to evaluate its effect in the B_s^0 channel. In the 2012 MC samples, the shifts in the $I_{0.7}$ variable total -8.9% in $\epsilon_{J/\psi\, K^+}$, and -3.8% in $\epsilon_{\mu^+\mu^-}$, with an effect of -5.3% in the efficiency ratio; the 2011 samples have a similar effect but with a smaller size. In total this systematic error on R_ϵ^k from $I_{0.7}$ alone is $\pm 3.2\%$. Other variables have absolute ratio shifts of less than 2% and absolute efficiency shifts of less than 5%, in either channel, and any correlations with $I_{0.7}$ are negligible. The total uncertainty per trigger category on R_ϵ^k is simply the sum in quadrature of the uncertainties over all discriminating variables - this is the dominant contribution to the systematic uncertainties shown in Table 5.4.

The third systematic source, reconstruction efficiency differences between B^+ and B_s^0 channels due to the kaon tracks and decay vertices, contributes $\pm 3.6\%$ to the total systematic uncertainty. The fourth and final systematic source, residual uncertainties in the trigger efficiencies, is evaluated through data-driven studies using $J/\psi\, K^+$ and $\mu^+\mu^-$ candidates. These residual uncertainties cancel, to a large extent, leading to an average uncertainty across the categories of $\pm 1.5\%$.

5.10 Signal Yield Extraction

Candidate events passing the baseline selection, the fake-BDT selection and the continuum-BDT selection are sorted into one of three intervals according to the continuum-BDT output: 0.240-0.346, 0.346-0.446 and 0.446-1. Each interval corresponds to a signal efficiency of 18%, and ordered in terms of increasing signal-to-background ratio. The trigger categories are combined in each interval, to maximise the number of events. An unbinned extended maximum likelihood fit is performed on the dimuon invariant mass distribution, simultaneously across the three continuum-BDT bins. The result of the fit gives the total yield of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ events. Any systematic uncertainties related to the BDT intervals and the models used to describe the signal and background are included in the likelihood as Gaussian multiplicative factors with width equal to the value of the systematic uncertainty.

Signal Model and BDT Interval Efficiency

The mass distribution for the $B^0 \to \mu^+\mu^-$ signal is described by a superposition of two Gaussian distributions, both centered at the B_d^0 or B_s^0 mass (5280 and 5366 MeV, respectively [3]) with the other parameters extracted from MC simulation. Systematic uncertainties from the mass scale and resolution are considered separately for the two models. The MC mass distributions with the double-Gaussian fits overlaid are shown in Figure 5.14, with the event yields normalised to the SM expectations.

Since the output of the continuum-BDT is used to define the fitting intervals, the efficiencies for both data and MC must be compatible across the three bins. As shown in Figure 5.12, there is a discrepancy in the B^+ reference channel between data and MC, in the BDT output distribution. This discrepancy has a linear dependence on the BDT output value, shown in the ratio plots in Figure 5.12, and the MC can be reweighted using that dependence. This dependence is equal within 12% between $B^+ \to J/\psi\, K^+$ and $B_s^0 \to J/\psi\, \phi$, and the mean value of the slope shown in the ratio plots is used to correct the BDT output distribution in the $B_s^0 \to \mu^+\mu^-$ MC sample. This corresponds to a change in the signal efficiency of +0.018 in the 1^{st} and -0.018 in the 3^{rd} BDT intervals, with no change in the 2^{nd} BDT interval.

Evaluating the systematic uncertainties for the BDT interval efficiency is done with a similar procedure as for the event selection, described in Section 5.9. The MC samples for the reference channels are reweighted according to the data-MC difference in each discriminating variable, and the systematic uncertainty due to variable mismodelling is found by taking the sum in quadrature of the individual changes in BDT interval efficiency after reweighting a single variable. These uncertainties are found

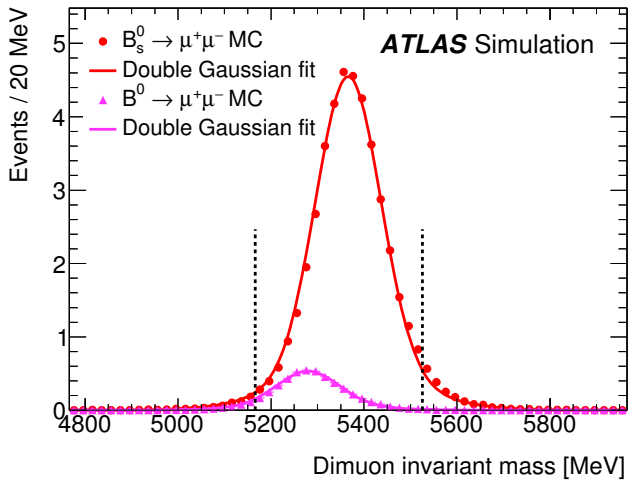

Figure 5.14: Dimuon invariant mass distributions for B_s^0 and B_d^0 signal decays, taken from MC samples, with the double-Gaussian fits overlaid. Both distributions are normalised to the SM expectation of the yield for an integrated luminosity of 25 fb^{-1}. This combines all three continuum-BDT intervals.

to be ± 0.026, ± 0.010 and ± 0.023 for the first, second and third intervals respectively. Gaussian terms are included in the likelihood in order to describe these uncertainties.

Figure 5.15 shows the continuum-BDT output for both reference channels, after the MC reweighting based on the BDT output has been applied, as well as the MC distribution for the $B_s^0 \to \mu^+\mu^-$ sample before and after reweighting, illustrating the systematic uncertainty discussed above. The isolation $I_{0.7}$ reweighting previously discussed in Section 5.9 is not applied to these distributions, or to the calculation of the BDT interval efficiencies. This is due to correlations between the discriminating variables being enhanced after the BDT selection is applied.

Background Model

Section 5.5 describes the three main components of the background, which are included in this fit. The combinatorial background is described with a first-order polynomial. According to simulation the slope of the continuum distribution is similar in the three BDT bins, with the correlation between BDT value and dimuon mass being small. Hence the slope of the mass dependence is described by independent parameters in the three intervals, subject to loose Gaussian constraints, allowing for up to 40% variation between the first and second intervals, and up to 80% between the first and third intervals. These variations are larger than were seen in simulation, but are consistent with those determined from data.

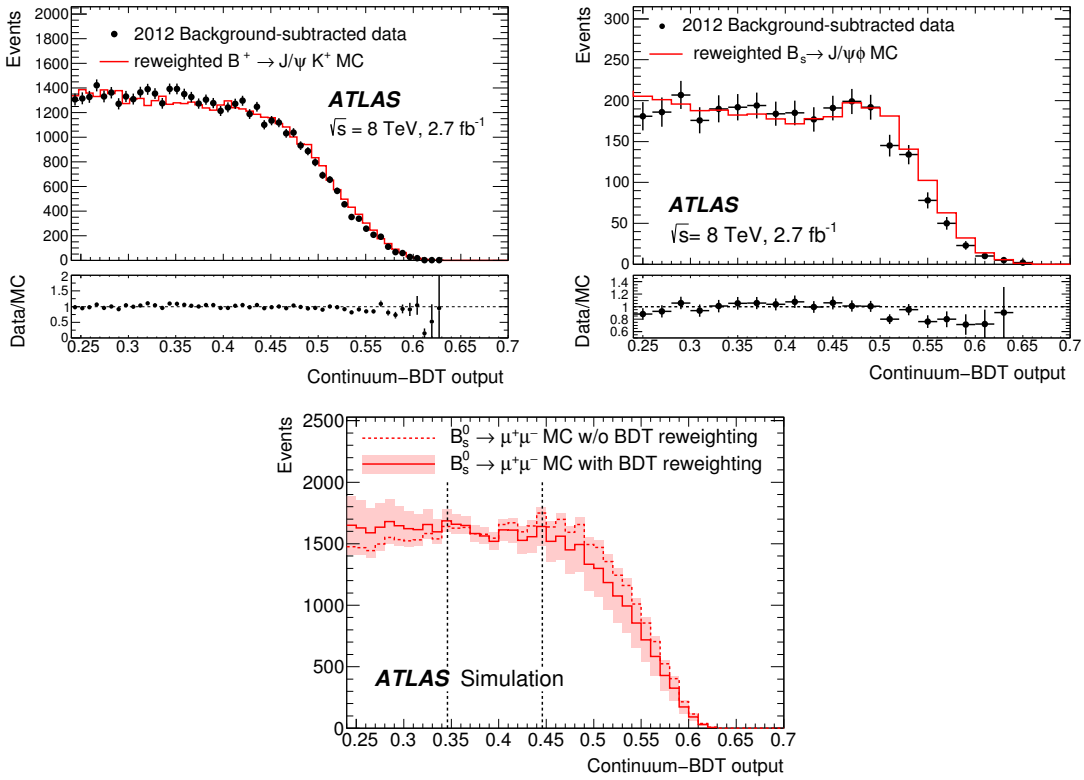

Figure 5.15: Background-subtracted data and MC distributions for the continuum-BDT output, in the $B^+ \to J/\psi\, K^+$ and $B_s^0 \to J/\psi\, \phi$ reference channels. The MC samples are normalised to the number of data events, with a linear correction applied based on the slope of the data-MC ratio in Figure 5.12. The lower plot shows the difference between the reweighted and not-reweighted distributions for the $B_s^0 \to \mu^+\mu^-$ MC sample, with the shaded area demonstrating the systematic uncertainty assigned to it. The vertical dashed lines correspond to the three BDT intervals.

The SS-SV background has a mass distribution that peaks below the low-mass region, the tail of which leaks into that sideband. The mass dependence of this background is derived from the signal-like low-mass sideband data, and is described by an exponential function with equal shape across all three BDT intervals. The value of the shape parameter is extracted from the fit, and the normalisation values are independent in each BDT interval. For the peaking background, which is very similar to the B^0 signal, a fixed mass shape is used, with a normalisation of 1.0 ± 0.4 events, as previously discussed in Section 5.6, distributed equally across the BDT intervals.

Pseudo-MC experiments are used to test and validate the fitting procedure, as discussed below. Using the three BDT intervals is found to optimise the performance of the likelihood fit - all three intervals

contribute to the determination of the partially-reconstructed decays, the first and second to the determination of the continuum background, and the second and particularly the third bins provide the sensitivity to the $B^0 \to \mu^+\mu^-$ signal.

Systematic Uncertainties in the Fit

A number of assumptions are used in the construction of the fit model, such as the uniformity of the combinatorial background mass dependence across the three BDT intervals. The sensitivity of the fit to these assumptions is assessed using pseudo-MC experiments, in which the descriptions of the signal and background components are varied. The corresponding deviations in the average numbers N_s, N_d of B_s^0 and B^0 events returned by the fit, with respect to the nominal configuration, are taken as systematic uncertainties.

The pseudo-MC experiments were generated with the background yield extracted from data in the mass sidebands. The signal itself was generated with different branching ratio configurations other than the nominal SM values, by varying the relative branching fraction ratio $\frac{\mathcal{B}(B_s^0)}{\mathcal{B}(B_d^0)}$.

For all variations generated, the pull distributions (the differences between results and generated values, divided by the fit errors) are found to be correctly described by Gaussian functions with widths approximately equal to one and mean values smaller than 0.2 (0.4) for B_s^0 (B_d^0). The fit biases are evaluated by performing the nominal fit on the pseudo-MC variations; for B_s^0, the fit bias is negligible, while for the B_d^0 the absolute value is smaller than 25% of the fit error and is included as an additional systematic uncertainty.

The variations in the number of fitted signal events, N_s or N_d, in the pseudo-MC samples are combined by taking the sum in quadrature of all positive or all negative shifts and using the larger of the two values as a symmetric systematic uncertainty. For B_s^0, the total systematic uncertainty is found to increase with the assumed size of the signal, corresponding to $\sigma_{syst}(N_s) = \sqrt{2^2 + (0.06 \times N_s)^2}$, while for the B_d^0, the total systematic uncertainty is approximately $\sigma_{syst}(N_d) = 3$. Most shifts observed have opposite signs for N_s and N_d, leading to a combined correlation coefficient of $\rho_{syst} = -0.7$. These systematic uncertainties are included by introducing two smearing parameters for N_s and N_d in the fit likelihood, which are constrained by a Gaussian distribution parameterised by $\sigma_{syst}(N_s)$, $\sigma_{syst}(N_d)$ and ρ_{syst}.

Results of the Signal Yield Extraction

Using the full 2011 and 2012 data, the number of background events in the signal region (5166-5526 MeV) is computed from the interpolation of data observed in the sidebands, in the three BDT inter-

vals. The signal region is then unblinded and the data within, a total of 1951 events in the full mass range (4766-5966 MeV), are fitted. The unblinded dimuon mass distributions are shown in Figure 5.16, while Table 5.6 shows the interpolated background expectations, the expected number of signal events according to the SM, and the number of fitted signal events. The fitted signal yields are obtained without applying any boundary on the values of the fitted parameters, and the uncertainties correspond to likelihood variations of $-2\Delta\ln(L) = 1$. The likelihood includes the systematic uncertainties previously discussed, but the statistical uncertainties largely dominate.

	continuum-BDT interval		
	0.240-0.346	0.346-0.446	0.446-1
Background	509 ± 28	32 ± 6	4.8 ± 1.9
Expected total N_s		41	
Expected total N_d		5	
Total fitted N_s yield		16 ± 12	
Total fitted N_d yield		-11 ± 9	

Table 5.6: Fit results including the expected background and signal yields in the 3 BDT intervals. Errors include systematic and statistical uncertainties. Event yields are not required to be positive, resulting in a negative B^0 yield.

The primary result of this analysis, detailed in the branching ratio extraction (Section 5.11), is obtained by requiring N_d to be non-negative. The same requirement is applied to the distributions in Figure 5.16). This results in a new fitted yield of $N_s = 11$ and $N_d = 0$. An alternate constraint can be applied for comparison, where N_d is constrained according to the SM expectation for the ratio $\mathcal{B}(B^0 \to \mu^+\mu^-)/\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ multiplied by the hadronisation probability ratio f_d/f_s [70] rather than being extracted from the fit simultaneously with N_s. This leads to a change in N_s by -0.8 to a value of $N_s = 10.2$, and in N_d by $N_d = N_s/8.3 \approx 1.2$.

5.11 Branching Ratio Extraction

With the signal yields found and all efficiencies evaluated, the branching fractions for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays can be extracted from data, using a profile-likelihood fit. The likelihood is similar to the one used to find N_s and N_d, replacing the fit parameters with the corresponding branching fractions divided by the normalisation terms from Eq. 5.1, and with Gaussian multiplicative factors for the normalisation uncertainties. The values of all input parameters can be found in Table 5.7.

The normalisation terms include the B^+ branching fraction and the relative hadronisation probability, both of which are taken from external sources rather than extracted from data. The B^+ branching fraction is obtained from world averages [3], as a product of $\mathcal{B}(B^+ \to J/\psi K^+) = (1.027 \pm 0.031) \times 10^{-3}$

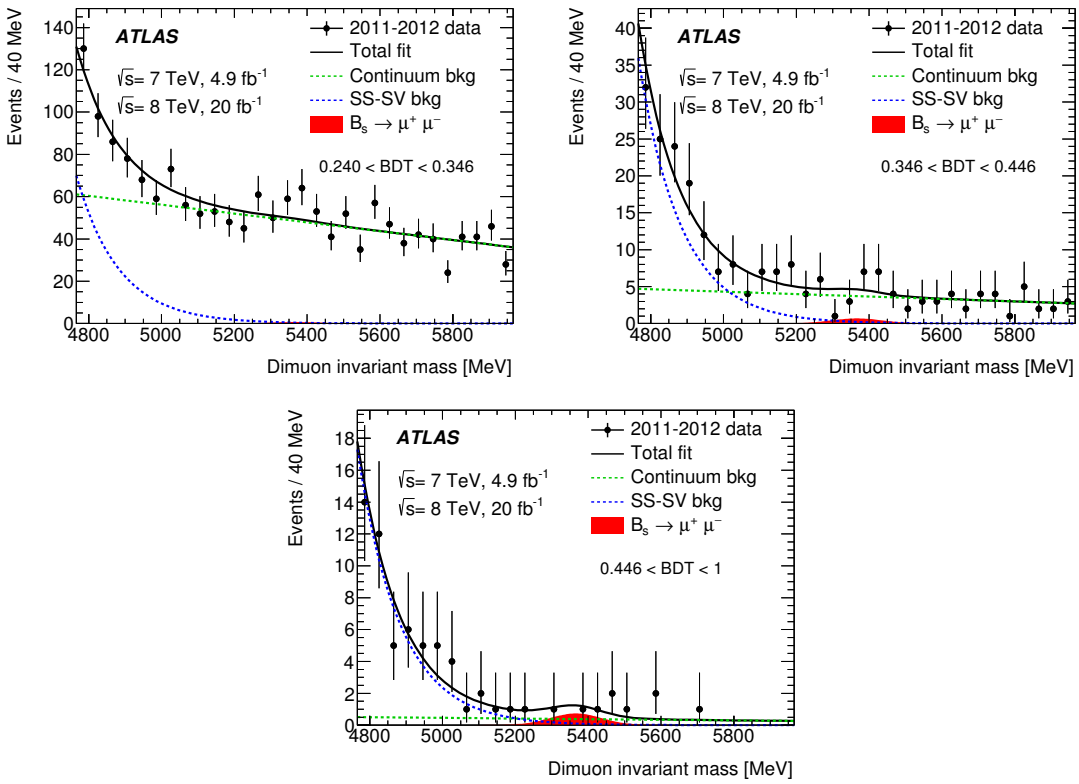

Figure 5.16: Dimuon invariant mass distributions in the unblinded data, in the three continuum-BDT intervals. The signal yields for both B_s^0 and B^0 are constrained to be non-negative. The total of the maximum-likelihood fit is shown as the black line, with the blue dashed line corresponding to the SS+SV background, the green dashed line the combinatorial background, and the red filled area the $B_s^0 \to \mu^+\mu^-$ signal.

and $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033)\%$. The hadronisation probability, $f_{s,d}/f_d$, is equal to 1 for B_d^0 and 0.240 ± 0.020 for B_s^0 as measured by ATLAS [70].

The number of events in the normalisation channel, weighted by efficiency and luminosity, enter into the $\mathcal{D}_{norm}$ term (Eq. 5.2) and are obtained from Tables 5.3 and 5.4 with the uncertainties listed in Table 5.5. Summing over each trigger category gives $\mathcal{D}_{norm} = (2.88 \pm 0.17) \times 10^6$ for B_s^0 and $(2.77 \pm 0.16) \times 10^6$ for B^0, with the different values arising due to the efficiency corrections applied independently to the two channels. Combining these values with the B^+ branching fraction and the hadronisation probabilities, discussed above, gives the single-event sensitivity, which is equal to $(8.9 \pm 1.0) \times 10^{-11}$ for B_s^0 and $(2.21 \pm 0.15) \times 10^{-11}$ for B_d^0 . If the requirement that the signal event yields are non-negative is applied, the value of the branching fractions that optimise the profile-

Parameter	Value
N_d	0
N_s	11
f_s/f_d	0.24 ± 0.02
$\mathcal{B}(B^+ \to J/\psi\, K^+)$	(5.961 ± 0.033) %
$N^k_{J/\psi K^+}$	T1: 46860 ± 290 ± 280
	T2: 5200 ± 84 ± 100
	T3: 2512 ± 91 ± 42
	2011: 95900 ± 420 ± 1100
R^k_ϵ	T1: 0.180 ± 0.001 ± 0.009
	T2: 0.226 ± 0.004 ± 0.014
	T3: 0.189 ± 0.005 ± 0.022
	2011: 0.156 ± 0.002 ± 0.009
$\mathcal{D}_{\rm norm}$ (B^0_s)	(2.88 ± 0.17) $\times 10^6$
$\mathcal{D}_{\rm norm}$ (B^0_d)	(2.77 ± 0.16) $\times 10^6$

Table 5.7: The values for each parameter that enters into the profile likelihood fit to extract the branching ratios for B^0_s and B^0_d .

likelihood fit are $\mathcal{B}(B^0_s \to \mu^+\mu^-) = 0.9 \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = 0$. These are the results used for the remainder of the section, i.e. the non-negative constraint is applied, unless otherwise stated.

A Neyman construction [71] is used to determine the 68.3% confidence interval for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ with pseudo-MC experiments, leading to the branching fraction:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = \left(0.9^{+1.1}_{-0.8}\right) \times 10^{-9}. \tag{5.3}$$

The uncertainties stated include both statistical and systematic contributions. The relative contributions can be found by repeating the likelihood fit with the total systematic uncertainty set to zero. The statistical uncertainty is dominant, with the systematic uncertainty equal to $\pm 0.3 \times 10^{-9}$.

Pseudo-MC experiments are used to determine the observed $B^0_s \to \mu^+\mu^-$ significance. A hypothesis test based on the likelihood ratio $-\ln[L(\text{no signal})/L(\text{max})]$ [72] gives a significance of 1.4 standard deviations for a SM signal. This compares to the expected significance for the SM $B^0_s \to \mu^+\mu^-$ branching fraction, calculated in the same manner, which is equal to 3.1 standard deviations. In this test, the $B^0 \to \mu^+\mu^-$ branching fraction is left free to be determined in the fit. The compatibility of this result with the SM is also evaluated with pseudo-MC experiments, and a hypothesis test based on

$-\ln[L(\text{SM})/L(\text{max})]$ [72] is performed for the simultaneous fit to the two signal branching ratios. The observed SM compatibility has a p-value (the probability that the data will fluctuate to the observed level assuming only the SM) of $p = 0.048 \pm 0.002$, corresponding to 2.0 standard deviations.

Given that the B_s^0 signal significance is not large enough to measure the branching ratio at 95% CL, an upper limit is set. Using the CL_s method [73], implemented with pseudo-MC experiments under a background-only hypothesis, and leaving the B_d^0 branching ratio free to be determined in the fit, an upper limit is placed on the $B_s^0 \to \mu^+\mu^-$ branching ratio at the 95% confidence level:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.0 \times 10^{-9} (95\%\text{CL}). \tag{5.4}$$

A similar procedure is used to set an upper limit on the $B^0 \to \mu^+\mu^-$ signal. In this case the pseudo-MC samples are generated according to the observed background and the B_s^0 signal. The observed limit is:

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 4.2 \times 10^{-10} (95\%\text{CL}). \tag{5.5}$$

These observed limits are above both the SM prediction and the central value of the measurement from CMS and LHCb [57], and thus there is no tension with theory or existing experimental studies.

The observed branching ratio, allowing for negative values of the $B^0 \to \mu^+\mu^-$ branching fraction, is shown in Figure 5.17, in the plane of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$. Contour lines have been drawn for values of $-2\Delta\ln(L) = 2.3$, 6.2 and 11.8 (corresponding to confidence intervals of 68.27%, 95.45% and 99.73% respectively) relative to the maximum of the likelihood. Also shown are the equivalent contour lines for the combined results of the CMS and LHCb experiments, as well as the SM prediction. The $B_s^0 \to \mu^+\mu^-$ branching fraction given in Eq. 5.3 is also shown, with error bars indicating the 68.3% confidence interval.

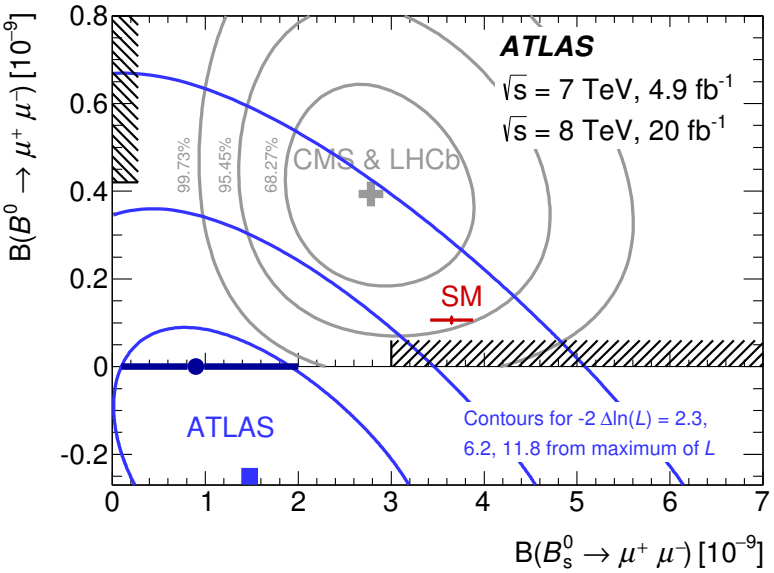

Figure 5.17: Contours in the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, $\mathcal{B}(B^0 \to \mu^+\mu^-)$ plane for intervals of $-2\Delta\ln(L)$ equal to 2.3, 6.2 and 11.8 relative to the absolute maximum of the likelihood. The branching fraction obtained by imposing the non-negative yields constraint, as per the text, is shown as the dark blue dot complete with error bars indicating the 68.3% confidence interval. The hatched areas indicate the upper limits on the B_s^0 and B^0 branching ratios at 95% CL. Contours from the combined CMS and LHCb result [57] are also shown, along with the SM prediction.

5.12 Conclusions

Rare B decays such as $B_s^0 \to \mu^+\mu^-$ are important indirect probes in the search for BSM physics. Their highly-suppressed nature makes them theoretically clean, meaning any deviations away from the SM are likely to be caused by new physics. However the extremely low branching fractions make them experimentally challenging to search for amongst an overwhelming background. This chapter has detailed a search for these rare decays, using the full 25 fb^{-1} dataset from the ATLAS detector, taken during 7 and 8 TeV proton-proton collisions in 2011 and 2012.

After observing a small number of signal events, an upper limit has been placed on the branching fractions of both $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. For the B_s^0 a measurement was done at the 68% CL, giving a result of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \left(0.9^{+1.1}_{-0.8}\right) \times 10^{-9}$ including both statistical and systematic uncertainties. An upper limit $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.0 \times 10^{-9}$ at 95% CL is placed, which is in agreement with the SM and the combined CMS and LHCb result, though is certainly lower than both. For the B^0, no events were observed, and an upper limit $\mathcal{B}(B^0 \to \mu^+\mu^-) < 4.2 \times 10^{-10}$ at 95% confidence level is set, which is compatible with both the SM prediction and the combined CMS and LHCb result.

A p-value of 4.8% (corresponding to 2 standard deviations) is found for the compatibility of the results with the SM prediction. This result does not rule out the presence of Supersymmetry, or BSM physics in general, but it does create strong constraints that any BSM theory must abide by.

Looking forward to the LHC Run 2 at $\sqrt{s} = 13$ TeV, searching for $B_s^0 \to \mu^+\mu^-$ decays at ATLAS becomes even more difficult. On the one hand, the increase in data (up to 100 fb^{-1} by the end of Run 2) could make up for the problem experienced here of a massive under-fluctuation of observed events, compared to the expectation. On the other hand, a higher centre-of-mass energy, combined with a smaller bunch spacing, will lead to more pileup. More pileup would require either an almost unworkable level of prescaling, or a muon p_T trigger threshold at 10 GeV or above - both of which would reject the majority of $B_s^0 \to \mu^+\mu^-$ events before the analysis even began, effectively negating the larger dataset. Regardless, the problem of the ATLAS muon mass resolution being too large to distinguish between B_s^0 and B_d^0 is still unsolved, and may not be solved until the LHC Run 3 or even the HL-LHC in the 2030s. In effect, while $B_s^0 \to \mu^+\mu^-$ searches at ATLAS in Run 2 and beyond are certainly possible, they may not be technically viable, and this result might well be the defining result for ATLAS for a decade.

Chapter 6

Direct Search for Hadronic Stop Pair Production

6.1 Introduction

Indirect probes for new physics, such as rare B-decays, are useful tools for finding evidence of any number of theories. However in the situation where there is a deviation away from the SM, only limited information about the underlying BSM theory can be extracted. The model-independent nature of these analyses is what makes them so powerful as a search tool, but as a trade-off, it is difficult to infer information about the underlying theory from an indirect search alone. Direct searches utilising specific, well-understood models are used instead to measure or to limit the parameters of BSM theories. These model-dependent searches may not be as generalised, but if a significant fluctuation away from the SM is observed, it can be easily interpreted theoretically.

Supersymmetry (SUSY) is an extension to the SM, that resolves the gauge hierarchy problem by introducing supersymmetric partner particles (sparticles) to the SM particles [18–20, 74–76]. More details on both SUSY and the hierarchy problem, as well as the full SUSY particle content, can be found in Chapter 2. While there are no strict predictions on the masses of the sparticles, large levels of fine-tuning can be avoided by requiring some to be relatively light, $\mathcal{O}(1\text{TeV})$. These 'naturalness' constraints place the lightest SUSY particles within the reach of the LHC [21].

Of particular interest to both experimentalists and theorists are the third-generation squarks, the stop ($\tilde{t}_1$) and sbottom ($\tilde{b}_1$), since the Yukawa coupling between top quarks, stops and the Higgs field is a driving force behind SUSY as a solution to the hierarchy problem. In addition, as the LHC is a proton-proton collider, where strong force interactions dominate, the squarks and gluinos ($\tilde{g}$) are predicted to have the highest production cross-sections. If SUSY exists and is natural, the third-generation squarks

are expected to be the lightest squarks ($m_{\tilde{q}} < 1$ TeV) and thus may be the first ones to be found. The gluinos do have a larger production cross-section than stops, at the LHC, however they are also heavier. Figure 6.1 shows the production cross-sections as a function of sparticle mass, simulated using NLO+NLL predictions - this shows that at $\sqrt{s} = 13$ TeV, a $m_{\tilde{t}_1} = 1000$ GeV stop has roughly the same cross-section as a $m_{\tilde{g}} = 1700$ GeV gluino.

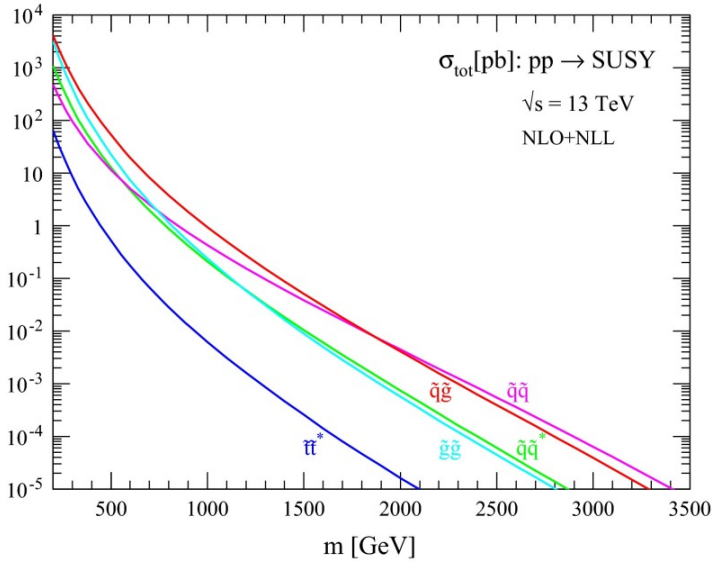

Figure 6.1: Production cross-sections at the LHC at $\sqrt{s} = 13$ TeV for the squarks and gluinos, using NLO+NLL predictions. [23]

In an R-parity conserving SUSY model, stops are produced in pairs through $q\bar{q}$ annihilation or gluon-gluon fusion [77, 78]. The two stop mass eigenstates, $\tilde{t}_1$ and $\tilde{t}_2$, are produced separately - mixed final states are highly suppressed, occurring only through electroweak interactions or QCD loop diagrams. The final step in any stop decay chain is, if R-parity is conserved, to the neutralino, $\tilde{\chi}_1^0$, predicted to be the lightest supersymmetric particle (LSP). The LSP is stable, as it can only decay to other supersymmetric particles, none of which are lighter than it; it is also neutrally-charged, which makes it impossible to detect directly using the LHC detectors, but provides an ideal candidate for dark matter [79–81].

This chapter presents a direct search for pair-production of hadronically decaying stops. This can occur through several possible decay channels, all of which are considered in this study:

- $\tilde{t}_1 \to t\tilde{\chi}_1^0$,

- $\tilde{t}_1 \to b\tilde{\chi}_1^{\pm} \to bW^{(*)}\tilde{\chi}_1^0$,
- or $\tilde{t}_1 \to bW\tilde{\chi}_1^0$ through a 3-body decay vertex.

These decay channels are shown as Feynman diagrams in Figure 6.2, interpreted as simplified models, described in Chapter 2. Stop decay final states are defined by the decay of the W, which can produce leptons or quarks. While the leptonic channel and the semi-leptonic channel - where one of the pair of stops decays leptonically and the other hadronically - have clean electron or muon signals to pick out of the background, the hadronic channel is swamped in QCD background. However, it also has the largest branching fraction of the three channels, at $> 50\%$ probability, which balances the high level of background. In addition, the lack of a neutrino in the final state means that any missing momentum comes solely from the neutralino.

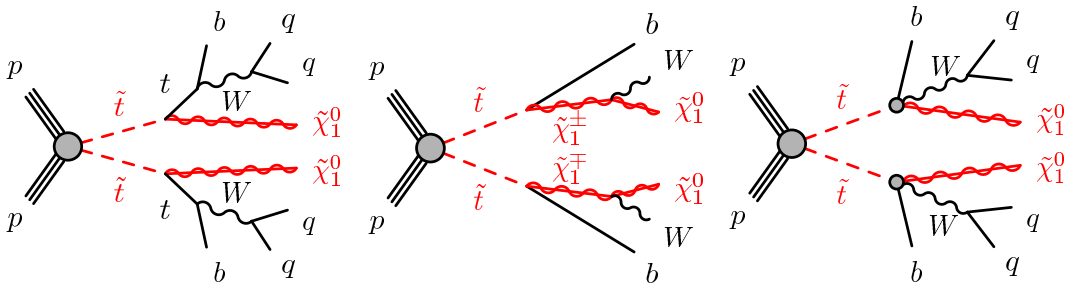

Figure 6.2: Feynman diagrams for the three stop decay topologies considered in this analysis. These simplified models contain a limited number of steps in the decay, with a specific topology involving four or more jets and missing transverse momentum. *From left to right:* $\tilde{t}_1 \to t\tilde{\chi}_1^0$, $\tilde{t}_1 \to b\tilde{\chi}_1^{\pm} \to bW^{(*)}\tilde{\chi}_1^0$, and $\tilde{t}_1 \to bW\tilde{\chi}_1^0$.

The nominal final state for the signal has a high jet multplicity, with at least four jets from the two b quarks and two W bosons, and large missing transverse momentum from the $\tilde{\chi}_1^0$ ($\mathbf{p}_T^{\text{miss}}$, the magnitude of which is the measured quantity, known as the missing energy E_T^{miss}). As such there are several background processes which can mimic the signal:

- $Z \to \nu\nu$ decays, produced in association with jets initiated by heavy-flavour quarks (the dominant background source),
- $W \to l\nu$ decays, where the lepton is missed or misreconstructed as a jet, produced in association with heavy flavour quarks,

- leptonic or semi-leptonic $t\bar{t}$ decays, where any leptons in the final state are missed or misreconstructed as jets,
- Diboson production, i.e. WW, WZ or ZZ ($Z \to \nu\nu$), where the W or one of the Z bosons decays hadronically and additional jets are produced by QCD interactions,
- $t\bar{t} + Z$, where the tops decay hadronically and the Z decays invisibly,
- and production of single top quarks in association with a W boson, for example by quark-gluon fusion, resulting in a semi-leptonic final state where the lepton is misreconstructed as a jet.

Fully hadronic top decays are, perhaps surprisingly, not a problem for a majority of the signal channels, as there is no mechanism for producing large amounts of missing energy. However, in compressed SUSY scenarios where the mass difference between the stop and the neutralino is small, $\mathcal{O}(m_t)$, the two stops tend to be produced back-to-back. The missing energy from the two neutralinos in the final state in this case are cancelled, resulting in an identical topology to non-resonant hadronic $t\bar{t}$ production. The technique for dealing with this involves selecting events with a high-p_T initial state radiation (ISR) jet which boosts the stop pair in one direction and creates E_T^{miss} again; a dedicated search region for this kind of decay is defined for this analysis.

The search is based on 36.07 fb^{-1} of pp collisions recorded by ATLAS in 2015 and 2016 at centre-of-mass energy $\sqrt{s} = 13$ TeV. A previous result with just 13.3 fb^{-1} luminosity was presented at the ICHEP 2016 conference, and a similar study was done during the LHC Run-1 at $\sqrt{s} = 8$ TeV [82, 83]. A search for sbottoms, the other third-generation squark, was performed by ATLAS in 2015 and 2016, and can be found in Ref. [84] - several parallels between that analysis and the one presented here can be made, especially in terms of analysis strategy and background modelling.

This chapter follows this outline: Section 6.2 covers the data and MC samples used in this analysis, including the trigger procedure used to record the data. Section 6.3 details the definitions of physics objects reconstructed in events. Section 6.4 describes the signal regions used to enhance sensitivity to $\tilde{t}_1$ decays, and Section 6.5 describes the control and validation regions used to extract the background yield estimations before looking at the signal regions. Section 6.6 covers the detector and theoretical systematic uncertainties included in this analysis. Finally, Section 6.7 contains the results of the analysis, including event yields in each signal region, p-values and new limits set.

While the bulk of my contributions to this analysis are contained in the next chapter, relating to reinterpretations and improvements for the nominal analysis described in this chapter, some work done personally by myself is shown here as well. Primarily, I contributed to the top quark reconstruction techniques used, and helped develop the (unused) Z background estimation method that uses a photon

as a proxy for the Z boson. In addition, I carried out several cross-checks for the validation regions described in Section 6.5, to verify the Monte Carlo estimates.

6.2 Data-taking and Monte Carlo Simulations

The data used in this analysis were collected during 2015 and 2016, with the LHC operating at $\sqrt{s} = 13$ TeV and 25 ns bunch spacing. It was during this time that the LHC reached its design instantaneous luminosity [85], resulting in the largest dataset yet recorded by ATLAS, with a total integrated luminosity of 36.07 fb^{-1}. A two-level trigger system is used to select events, as opposed to the three-level system used during the LHC Run 1. The Level 1 trigger is hardware-based and uses a subset of the detector information to reduce the accepted data rate to a maximum of 100 kHz. The second stage is the HLT, a software-based trigger that uses the full detector reconstruction to reduce the event rate to 1 kHz for offline storage.

The primary trigger used to collect the data used in this analysis is a E_T^{miss} trigger. In 2015, with a lower instantaneous luminosity, the trigger reconstructed the missing energy by doing a vector sum of all calorimeter cells, and required an event to have $E_T^{\text{miss}} > 70$ GeV to avoid requiring prescaling. The increased peak luminosity in 2016 led to the need to improve the trigger algorithm. The resulting trigger reconstruction algorithm is based on the transverse vector sum of all reconstructed jets, with a higher energy threshold, unprescaled at 110 GeV and above. Both the 2015 and 2016 triggers are fully efficient for events with $E_T^{\text{miss}} > 250$ GeV, verified using simulated MC signal samples.

For background studies where leptons are present in the final state, data were collected with electron or muon triggers, as a E_T^{miss} trigger would not be efficient. The electron trigger selects events based on the presence of clusters of energy in the electromagnetic calorimeter, with a shower shape consistent with that of an electron and a matching track in the ID; the muon trigger selects events with one or more muon candidates based on tracks in the MS and ID. For 2015, the minimum transverse momentum requirements were 24 GeV for electrons and 20 GeV for muons, while in 2016 this was increased to 26 GeV for both and a tight isolation requirement added, as a result of the higher instantaneous luminosity. In order to recover efficiency for high-p_T leptons, which may be lost due to the tight isolation requirement, additional single-lepton triggers with looser or no isolation requirements were also used, but with higher p_T thresholds. For electrons these additional triggers required $p_T > 60$ or $p_T > 120$ (140) in 2015 (2016), depending on the looseness of the isolation requirement. The additional muon triggers required $p_T > 50$ in both 2015 and 2016, with no requirement on the muon isolation.

Finally, several triggers based on the presence of high-p_T jets are used to collect data for the estimation of multijet and all-hadronic $t\bar{t}$ backgrounds. The jet p_T thresholds of these triggers range from 20 to

400 GeV - some of these triggers are prescaled, in order to reduce the data-taking rate contributed by the lower-p_T triggers to stay within the bandwidth limits of the trigger system.

Monte Carlo Samples

Simulated events are used to model the SUSY signal, as per the simplified models shown in Figure 6.2, and to aid the description and estimation of the background. The different signal channels are defined not just in terms of the decay products, but also the mass differences (Δm) between the sparticles:

- $\tilde{t}_1 \to t\tilde{\chi}_1^0$, with $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0) > m_t$, where m_t is the mass of the top quark;
- $\tilde{t}_1 \to b\tilde{\chi}_1^\pm \to bW^{(*)}\tilde{\chi}_1^0$, where $m(\tilde{\chi}_1^\pm) = 2m(\tilde{\chi}_1^0)$ as motivated by gaugino universality;
- and the three-body decay $\tilde{t}_1 \to bW\tilde{\chi}_1^0$, where $m(b) + m(W) < \Delta m(\tilde{t}_1, \tilde{\chi}_1^0) < m_t$.

Signal samples are generated in a grid across the $m_{\tilde{t}_1}$-$m_{\tilde{\chi}_1^0}$ plane, varying both masses in steps of 50 GeV, as the kinematics of decays even within the same channel may differ greatly depending on the mass of the sparticles. Benchmark models used for signal region optimisation are generated with $100k$ events each, while other mass points are generated with 10-20k events.

All signal models are generated using MadGraph5_AMC@NLO [86], coupled with PYTHIA 8 [60] for the parton showering (PS) and hadronisation and with EvtGen [61] to model b-decays. The matrix-element (ME) calculations are performed at tree-level, and include the emission of up to two additional partons. Signal cross-sections are calculated to next-to-leading order and next-to-leading-logarithmic accuracy (NLO+NLL) [77, 78] - the nominal cross sections and the uncertainties are taken from an envelope of cross-section predictions using different parton distribution function sets, factorisation and renormalisation scales [87]. Figure 6.3 shows examples of the production diagrams simulated for the stop signal.

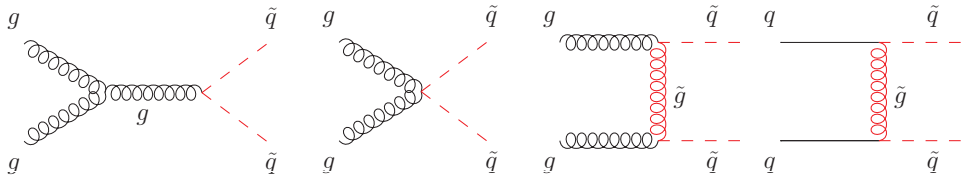

Figure 6.3: Example Feynman diagrams for tree-level squark pair production. These diagrams enter the matrix-element calculation during the signal simulation.

The SM background samples are simulated using different MC generators, depending on the process. Vector-boson production in association with final-state jets (Z + jets, W + jets) is generated with SHERPAv2.2.1 [88], in non-overlapping slices of boson p_T and filtered by jet flavour. Both semi-leptonic $t\bar{t}$ production and single top production are modelled with Powheg-Box v.2 [89] and interfaced to PYTHIA 6 [59] for PS and hadronisation, while $t\bar{t}$ + Z events are simulated with MadGraph5_AMC@NLO and PYTHIA 8. Finally the diboson samples are generated at leading order with SHERPAv2.2 - all other background sources are generated at next-to-leading order.

The ATLAS detector response is simulated using GEANT4 [62], or a fast-simulation framework whereby the calorimeter showers are simulated with a parametrised description and the rest is done by GEANT4. The effect of pileup is simulated by overlaying numerous minimum-bias interactions on top of the hard-scattering event. The simulated events are then reweighted in order to match the distribution of the average number of interactions per bunch crossing in data. Further reweighting to the simulated events is applied to correct for differences in the lepton trigger and reconstruction efficiencies, momentum scale, energy resolution, isolation and b-jet identification efficiency (and misidentification probability of light- and charm-jets); all of these reweightings are included as systematic uncertainties.

6.3 Event and Physics Object Reconstruction

Events that pass the trigger and are made available for physics, are required to have a PV [90]. A PV is a point of interaction between colliding protons - a single event may have many PVs. A PV is reconstructed from at least two associated tracks in the ID with $p_T >$ 400 MeV which are compatible with originating from the luminous region. In the case where multiple vertices are found, the PV is defined as the vertex with the largest summed p_T^2 of the associated tracks.

Jet Definitions

Jets are reconstructed from neighbouring clusters of energy deposits in the hadronic calorimeter using the anti-k_t jet clustering algorithm [91] with a distance parameter $R = 0.4$ and full four-momentum recombination. A correction is applied to account for additional energy deposited in the calorimeter cells by unrelated particles, according to an estimate of the pileup activity in a given event [92]. These calibrated jets are required to have $p_T > 20$ GeV, though no η requirement is applied at this stage - jets passing this cut ('baseline' jets) are used for overlap removal (see the electron and muon definitions below), and for determining whether an event should be rejected due to a large number of poorly-reconstructed jets [93]. Additional jets arising from non-collision sources (cosmic-ray showers, or proton losses from upstream of the interaction point) or detector noise are removed

from consideration [93]. After these quality selections, a pseudo-rapidity requirement is applied, only accepting jets with $|\eta| < 2.8$. An additional selection based on the output of a jet vertex-tagging algorithm is applied to jets with $|\eta| < 2.4$ and $p_T < 60$ GeV, to reject pileup jets not coming from the primary vertex. Jets passing all of these criteria are called 'signal' jets and are considered in this analysis.

A subset of the baseline jets passing the selection above are identified as originating from a b-quark decay by using a multivariate jet tagger algorithm [94], if they are within the ID acceptance ($|\eta| < 2.5$). This algorithm exploits the impact parameters of the tracks within the jet, the presence of secondary vertices and the reconstructed flight paths of b- and c- hadrons inside the jet. A working point was chosen for this tagger that gives an average 77% efficiency for identifying jets containing b-quarks, as measured with simulated $t\bar{t}$ events. Non-b jets are rejected by a factor of approximately 134 for light-quark and gluon jets (depending on the jet p_T), and 6.2 for charm jets [94, 95].

Electron Definition

Electrons are reconstructed from energy clusters in the electromagnetic calorimeter, that can be matched to a track in the ID. The pseudo-rapidity of the electron is required to be within the range $|\eta| < 2.47$, with $p_T > 7$ GeV; a likelihood-based selection is also applied, with a very loose working point. Candidate electrons that fall into the 'crack' region of the detector, $1.37 < |\eta| < 1.52$, but pass this likelihood-based selection, are also accepted as electrons. As an electron may be inside a jet, given that the jet reconstruction is independent of the electron reconstruction, an overlap removal algorithm is applied. If the separation between electron candidate and a non-b-tagged jet is $\Delta R < 0.2$, the object is considered as an electron - unless the jet is b-tagged, in which case the object is considered to be a b-tagged jet. If the separation between the two is within $0.2 < \Delta R < 0.4$, the object is defined as a jet.

To be considered a signal electron, a baseline electron must pass the additional signal selection. Signal electron candidates are required to have $p_T > 20$ (28) GeV for regions using a E_T^{miss} (lepton) trigger, and to satisfy p_T-dependent track- and calorimeter-based isolation criteria. The candidate must be matched to a trigger electron candidate, if using a lepton trigger, with $p_T > 2$ GeV over the trigger p_T threshold. A 'tight' likelihood-based selection is applied, to match the shape of the shower in the electromagnetic calorimeter to the expected shape for an electron. The impact parameter of the electron in the transverse plane with respect to the PV ($|d_0|$) is required to be less than five times the impact parameter uncertainty (σ_{d_0}); the impact parameter along the beam direction, $|z_0 \times \sin\theta|$, is required to be < 0.5 mm.

Muon Definition

The muon definition is similar to the electron definition, though instead of energy clusters in the calorimeters, matching tracks in the MS and ID are used to define the muon. These tracks must have $|\eta| < 2.7$ and $p_T > 6$ GeV, and must pass a likelihood-based selection at a loose working point to be matched to expected track shapes for a muon. Overlap removal is also applied to muons, though it is simpler than for electrons - if the separation between a muon candidate and a jet is $\Delta R < 0.4$, the object is considered to be a jet rather than a muon.

Baseline muons passing the following selection are called signal muons. The muon p_T must be > 20 (28) GeV for regions using a E_T^{miss} (lepton) trigger, and to satisfy p_T-dependent track- and calorimeter-based isolation criteria. Similarly to the electrons, the muon must be matched to a trigger muon candidate as appropriate, and have $p_T > 2$ GeV above the relevant lepton trigger p_T threshold. The muon must pass a 'medium' quality selection to clean the event of cosmic-ray muons, muons from proton radiation upstream of the interaction point, or fake muons from high-energy hadrons that have punched-through the calorimeter into the MS [96]. Impact parameter criteria are imposed, requiring $|d_0| < 3\sigma_{d0}$ and $|z_0 \times \sin\theta| < 0.5$ mm.

Tau Identification

Given that taus decay to a neutrino and a $W^{(*)}$ inside the detector volume, and the W may decay hadronically, it is necessary to be able to differentiate between a tau-jet and a jet in order to fully veto leptons. A jet is identified as a τ candidate if the jet is the closest jet to the E_T^{miss} with $\Delta\phi(E_T^{\text{miss}}, jet) < \pi/5$ and four or less tracks are associated with the jet. Only non-b-tagged jets are considered for τ candidacy, and thus jets with $|\eta| > 2.5$ are not considered. This τ identification was found to be effective during Run 1 - studies of the systematic uncertainty associated to the τ veto performed in the Run 1 analysis found the uncertainty to be negligible [83], and this is assumed to still be the case in Run2.

Missing Energy Calculation

The $\mathbf{p}_T^{\text{miss}}$ is the negative vector sum of the p_T of all baseline physics objects in the event, as defined above. An additional term is added to account for small energy depositions in the event that aren't associated to any selected object, calculated from ID tracks with $p_T > 400$ MeV matched to the PV in order to make it resilient to pileup contaminations. The magnitude of the vector is called the missing transverse energy, E_T^{miss}. The missing transverse momentum from the tracking system alone (with

magnitude $E_T^{\text{miss,track}}$) is computed from the vector sum of the reconstructed ID tracks with $p_T > 500$ MeV, $|\eta| < 2.5$, that are associated to the PV.

Top Quark Reconstruction

Not only is the LHC a top-quark factory, producing millions of tops every year, but the signal studied in this analysis has two tops in the final state. Being able to reconstruct these tops, to directly measure their properties, is a very powerful tool. For signal regions targeting the fully resolved $\tilde{t}_1 \to t\tilde{\chi}_1^0$ decay, reconstructing two hadronic top quarks helps reduce the QCD, Z+jets, W+jets and leptonic $t\bar{t}$ backgrounds. This is done by reclustering the baseline jets in the event into large $R = 1.2$ jets using the anti-k_t jet algorithm, joining neighbouring $R = 0.4$ jets together to make the larger jets. The reclustered jet with the highest-p_T is defined as the leading top candidate, and the second-highest-p_T reclustered jet the subleading top candidate. Additionally, an alternate reclustering is done with $R = 0.8$ as the distance parameter, in the case of highly-boosted tops where the decay products are more colinear. The masses of these reclustered jets are denoted as $m^0_{\text{jet},R=1.2}$, $m^1_{\text{jet},R=1.2}$, $m^0_{\text{jet},R=0.8}$and $m^1_{\text{jet},R=0.8}$, with the 0 and 1 indexes representing the leading and sub-leading candidates respectively.

Five categories of top reconstruction are defined, based on the masses of the two leading $R = 1.2$ reclustered jets, as listed below. These categories are labelled by what the reclustered jet could be, depending on its mass - **T**, for a reconstructed top, **W**, for a reconstructed W boson, and **0**, for when the reclustering fails to reconstruct either decay product. Assuming that every signal event includes two tops from the decays of the pair-produced $\tilde{t}_1$s, this failure of the reconstruction could be due to low-p_T top quarks producing decay products that are widely separated, or non-resonant $t\bar{t}$ production giving very soft jets. As an example, the **TW** category has one high-mass reclustered jet and one medium-mass reclustered jet, indicative of an event with a single top quark and a W boson, or a $t\bar{t}$ event where one quark from the sub-leading top decay has been missed.

- W0: $60 < m^0_{\text{jet},R=1.2} < 120$ GeV, $m^1_{\text{jet},R=1.2} < 60$ GeV,
- T0: $m^0_{\text{jet},R=1.2} > 120$ GeV, $m^1_{\text{jet},R=1.2} < 60$ GeV,
- WW: $60 < m^0_{\text{jet},R=1.2} < 120$ GeV, $60 < m^1_{\text{jet},R=1.2} < 120$ GeV,
- TW: $m^0_{\text{jet},R=1.2} > 120$ GeV, $60 < m^1_{\text{jet},R=1.2} < 120$ GeV,
- TT: $m^0_{\text{jet},R=1.2} > 120$ GeV, $m^1_{\text{jet},R=1.2} > 120$ GeV.

The objective of this categorisation is two-fold: to attempt to recover the sub-leading top candidate (or at least a W candidate) when the reclustering fails in the T0 and W0 categories, and to separate

different event topologies to optimise signal region cuts independently. Section 6.4 details the optimisation done for two signal regions that utilise this top reconstruction categorisation to maximise significance.

Figure 6.4 shows the percentage of events in each reconstruction category, for two MC signal samples. The largest fraction of events falls into the T0 category - where the reclustering has failed to find either a second top or W. In the Run 1 analysis, these events would have been effectively wasted despite being potentially significant. In addition, the T0 category contains a larger fraction of the background compared to the other categories, as few background sources can produce a large number of events with two heavy reclustered jets. An independent optimisation in this category will maximise significance for those particular events, however the ideal solution would be to improve the top reconstruction to move some of these events (again, assuming two tops are present in every signal event) from T0 into the TW or even TT categories, where there is a much lower background contribution.

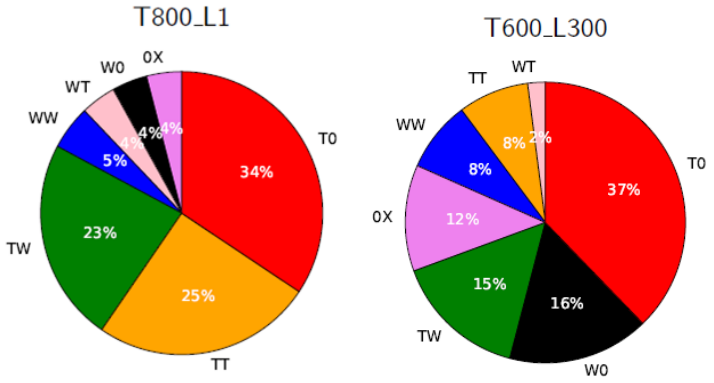

Figure 6.4: Percentage of simulated $\tilde{t}_1 \to t\tilde{\chi}_1^0$ events in each top reconstruction category as defined above, for two signal samples. *Left*: $m_{\tilde{t}_1} = 800$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV, a signal that would be targeted by SRA. *Right*: $m_{\tilde{t}_1} = 600$ GeV, $m_{\tilde{\chi}_1^0} = 300$ GeV, targeted by SRB. The jets used as input to the reclustering done here use the Run 1 jet definition, described in Ref. [83] - the notable difference is a tighter p_T cut, requiring $p_T > 35$ GeV.

In the Run 1 analysis, the definition of a baseline jet required the p_T to be greater than 35 GeV. This was due to the efficiency of the jet vertex finder algorithm used at the time, which reached peak efficiency at jet-p_T values of 35 GeV or higher. The tighter p_T cut also helped reduce Z + jets background at an early stage in the analysis. However, as the reclustering algorithm used to reconstruct tops relies on the baseline jets rather than the signal jets, if multiple soft, low-p_T jets radiated off harder jets are missed by the algorithm, it may have a large effect on the resulting top candidate mass. This is confirmed by the truth-level study shown in Figure 6.5. Truth-level partons are clustered into 'parton jets' with $R = 0.4$, then again into $R = 1.2$ top candidate jets, in the T0 and

TT categories. A p_T cut is then applied on the parton jets, and the reclustering algorithm ran again to form the top candidates. As the distributions show, even at truth level the reclustering struggles to reconstruct the top when low-p_T jets are missing from the input.

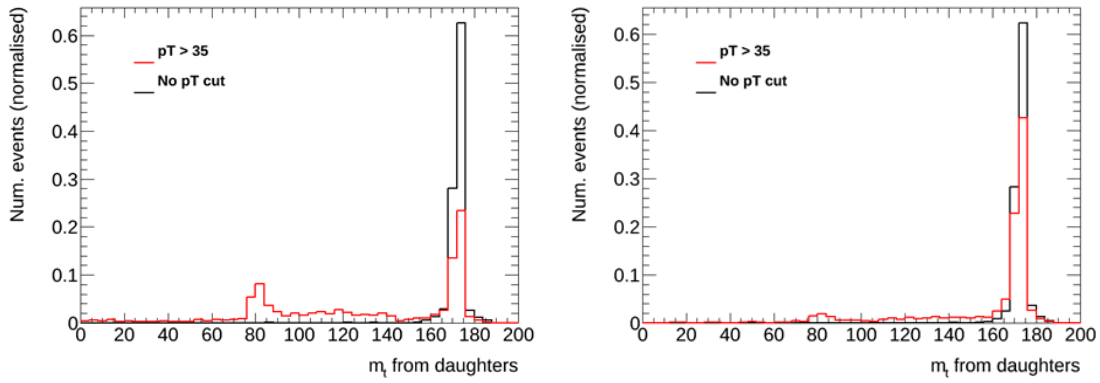

Figure 6.5: Truth-level partons are clustered into $R = 0.4$ jets, then the jets reclustered into $R = 1.2$ top candidates. This is done both with a p_T cut on the $R = 0.4$ jets (shown in red) and without (in black). The mass distributions of the reclustered top candidates is shown for *left*: the T0 category, and *right*: the TT category. The histogram areas have been normalised to unity to compare the distribution shape.

Figure 6.6 shows the effect of lowering the jet-p_T cut from 35 GeV to 20 GeV. This test was done using MC events that have gone through ATLAS detector simulation, using the same baseline jet definition described previously. The baseline $R = 0.4$ jets are 'matched' to the parton jets by requiring the separation in pseudorapidity between the two to be $\Delta R < 0.4$. This reduces the sensitivity of this study to pileup jets, ensuring that the reclustered jets are formed from only the signal decay products. The leading and subleading top candidate jets are shown, both of which are required to contain at least two jets that have been matched to parton jets. As can be seen, when the jet-p_T is required to be > 35 GeV, the subleading top candidate jet in the T0 category completely fails to peak at either the top mass or the W mass - even in the TT category, the subleading top candidate mass has a significant low-mass tail, though there is a clear peak. By reducing the jet p_T cut to 20 GeV, a clear peak is formed at the top mass in the subleading top candidate distribution, even in the T0 category.

This top reconstruction study was a driving force for changing the definition of a baseline jet to $p_T > 20$ GeV for Run 2 - the initial increase in Z+jets background, which is dominated by softer jets, is dealt with by maintaining a requirement that there are four or more jets in the event with $p_T > 40$ GeV. Further studies into alternative top reconstruction algorithms were completed; the reclustering algorithm was found to still perform best.

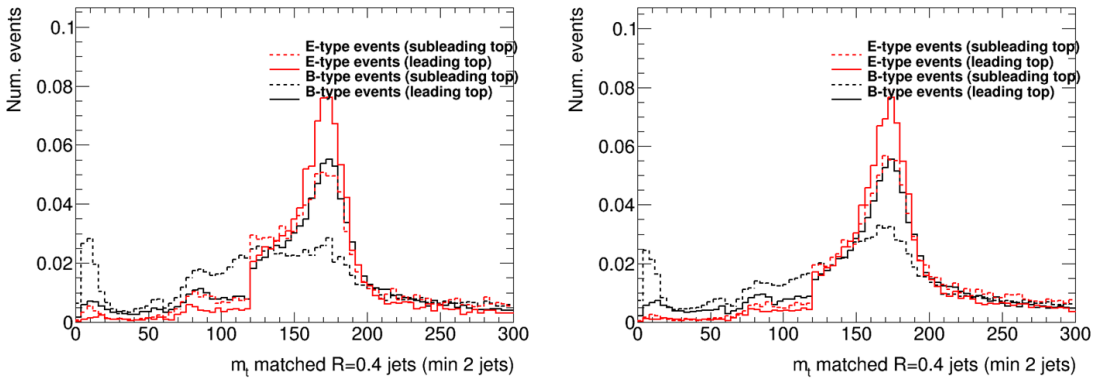

Figure 6.6: Mass distributions of the reclustered top candidate jets, in the TT (red) and T0 (black) categories. Leading top candidates are shown as solid lines, and subleading candidates as dashed lines. *Left*: a $p_T > 35$ GeV cut is applied to all baseline $R = 0.4$ jets, before they are used as input to the top reclustering. *Right*: the jet-p_T cut is reduced to 20 GeV. The effect of soft, low-p_T jets on the top reclustering is obvious when looking at the subleading top candidates; the 20 GeV cut results in a peak at the top mass in all cases, with small low-mass tails, whereas with a 35 GeV jet-p_T cut, the subleading candidate does not clearly peak, especially in the T0 category. The histogram areas have been normalised to unity to allow a comparison of shape rather than event yields.

6.4 Signal Region Definitions

All three signal processes share a similar event topology, namely multiple jets of which some originate from b-quarks, no leptons and large missing transverse momentum. A common preselection can be defined for all signal regions that exploits this common topology to reduce the background before any signal region optimisation is done. Table 6.1 describes this preselection, which is applied to all signal and validation regions (except for the 1-lepton W validation region); control regions have a separate preselection, described in Section 6.5, due to the presence of leptons. The leading four jets must satisfy $p_T > 80, 80, 40, 40$ GeV, and any events containing electrons or muons are rejected. The use of the E_T^{miss} trigger, as described in Section 6.2, dictates the requirement $E_T^{\text{miss}} > 250$ GeV, which also rejects the majority of background from QCD multijet and all-hadronic $t\bar{t}$ events. To further reduce multijet backgrounds, the angular separation between the azimuthal angle of the two highest p_T jets and the E_T^{miss} is required to satisfy $|\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})| > 0.4$, which rejects events with mis-measured E_T^{miss} arising from jet mis-calibration. The $E_T^{\text{miss,track}}$ must also be aligned in ϕ with respect to the E_T^{miss} in order to reduce E_T^{miss} mismeasurement, leading to the requirements $E_T^{\text{miss,track}} > 30$ GeV and $|\Delta\phi(E_T^{\text{miss}}, E_T^{\text{miss,track}})| < \pi/3$ radians.

This analysis defines four signal regions, each targeting a specific topology or kinematic regime of the

Trigger	Data 2015: `HLT_xe70_tc_lcw_L1XE50` Data 2016 (April - June): `HLT_xe90_mht_L1XE50` Data 2016 (June - July): `HLT_xe100_mht_L1XE50` Data 2016 (July onwards): `HLT_xe110_mht_L1XE50`
E_T^{miss}	> 250 GeV
N_{lep}	0
anti-k_t $R = 0.4$ jets	≥ 4, $p_T > 80, 80, 40, 40$ GeV
b-tagged jets	≥ 2
$\lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert$	> 0.4
$E_T^{\text{miss,track}}$	> 30 GeV
$\lvert\Delta\phi(E_T^{\text{miss}}, E_T^{\text{miss,track}})\rvert$	$< \pi/3$

Table 6.1: Preselection criteria used in all signal regions and the hadronic validation regions.

signal models. SRA and SRB are sensitive to production of high-mass $\tilde{t}_1$ pairs through the $\tilde{t}_1 \to t\tilde{\chi}_1^0$ decay chain - SRA targets signal with a large mass splitting between stop and neutralino ($\Delta\, m(\tilde{t}_1, \tilde{\chi}_1^0) \gg m_t, \sim 800-1000$ GeV), and SRB targets signal with a smaller mass splitting ($\Delta\, m(\tilde{t}_1, \tilde{\chi}_1^0) > m_t, \sim 400$ GeV). SRC is designed for the highly-compressed region where $\Delta\, m(\tilde{t}_1, \tilde{\chi}_1^0) \sim m_t$, using initial state radiation to improve sensitivity to these decays by boosting the $\tilde{t}_1$ pair in the transverse plane. Finally, SRD targets the two-step decay $\tilde{t}_1 \to b\tilde{\chi}_1^\pm \to bW^{(*)}\tilde{\chi}_1^0$, where no top quark candidates are reconstructed.

Signal Regions A and B

Given that SRA and SRB both target the same decay process, where top squarks decay via either $\tilde{t}_1 \to t\tilde{\chi}_1^0$ or the three-body decay $\tilde{t}_1 \to bW\tilde{\chi}_1^0$ with large mass splitting $\Delta\, m(\tilde{t}_1, \tilde{\chi}_1^0) > m_t$, both signal regions share similar techniques and selection criteria. Table 6.2 describes the full set of cuts required by both signal regions. As well as passing the preselection criteria, both regions require at least 2 b-tagged jets and 2 reclustered $R = 1.2$ top candidate jets. The top reconstruction categories, described in the previous section, are used here - the leading top candidate must be a well-reconstructed top, and events are then sorted based on the subleading candidate mass and which daughter particle is reconstructed. These categories are individually optimised due to the different signal-to-background ratio in each category. The three categories are statistically orthogonal, which allows them to be combined to maximise significance.

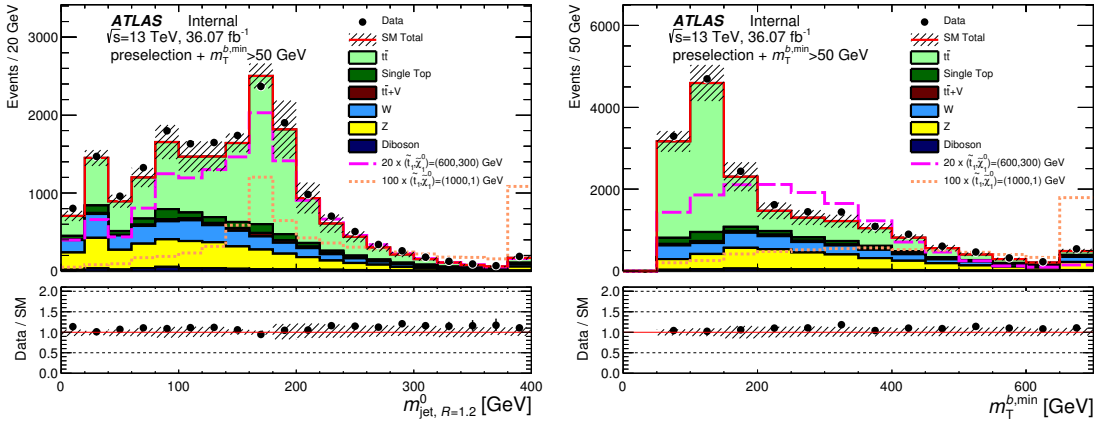

Figure 6.7: Distributions of two discriminating variables ($m^0_{\text{jet},R=1.2}$, left, and $m_T^{b,\text{min}}$, right) with 36.07 fb^{-1} of data, after the common preselection cuts and an additional $m_T^{b,\text{min}} > 50$ GeV cut are applied. The stacked histograms show the SM background expectations, and the dashed lines the prediction for two signal samples, $(m_{\tilde{t}_1}, m_{\tilde{\chi}^0_1}) = (800,300)$ and $(1000,1)$ GeV, x20 and x100 respectively to make them visible. The hatched error band around the SM expectation includes both statistical and systematic sources of uncertainty, and the right-most bin contains all events above the range of the histogram.

The most powerful variable in these two signal regions, for discriminating against hadronic $t\bar{t}$ production, is the E_T^{miss}. In order to reject (semi-)leptonic $t\bar{t}$ that produces missing energy through neutrinos, a requirement on $m_T^{b,\text{min}}$ is applied - the transverse mass (m_t) calculated from the E_T^{miss} and the b-tagged jet closest in ϕ to the $\mathbf{p}_T^{\text{miss}}$ direction must be above 200 GeV:

$$m_T^{b,\text{min}} = \sqrt{2p_T^b E_T^{\text{miss}} \left[1 - \cos\Delta\phi\left(\mathbf{p}_T^b, \mathbf{p}_T^{\text{miss}}\right)\right]} > 200\text{GeV}. \tag{6.1}$$

The distribution of $m_T^{b,\text{min}}$ is shown in Figure 6.7 after the common preselection cuts (and an additional $m_T^{b,\text{min}} > 50$ GeV cut) are applied. As the distribution shows, this cut at 200 GeV dramatically reduces the $t\bar{t}$ background without significantly impacting the signal.

A τ veto is applied to both signal regions, whereby events containing a τ as identified by the criterion described in Section 6.3 are rejected. Additional discrimination in SRB is provided by $m_T^{b,\text{max}}$, which is analogous to $m_T^{b,\text{min}}$ except the mass is calculated using the b-tagged jet furthest in ϕ from the $\mathbf{p}_T^{\text{miss}}$ direction, and $\Delta R(b,b)$, the angular separation between the two jets with the highest b-tagging weights.

Finally, additional background reduction is achieved in SRA by utilising m_{T2}, defined as the transverse mass between the subleading-p_T jet and the E_T^{miss}:

Signal Region		TT	TW	T0
	$m^0_{\text{jet},R=1.2}$	> 120 GeV		
	$m^1_{\text{jet},R=1.2}$	> 120 GeV	$60 - 120$ GeV	< 60 GeV
	$m_T^{b,\min}$	> 200 GeV		
	b-tagged jets	≥ 2		
	τ-veto	yes		
A	$m^0_{\text{jet},R=0.8}$	> 60 GeV		
	$\Delta R(b,b)$	> 1	-	
	m_{T2}	> 400 GeV	> 400 GeV	> 500 GeV
	E_T^{miss}	> 400 GeV	> 500 GeV	> 550 GeV
B	$m_T^{b,\max}$	> 200 GeV		
	$\Delta R(b,b)$	> 1.2		

Table 6.2: Selection criteria for SRA and SRB, in addition to the common preselection requirements described in the text. The signal regions are separated into topological categories based on reconstructed top candidate masses.

$$m_{T2} = \sqrt{2p_T^{\text{jet2}} E_T^{\text{miss}} \left(1 - \cos\Delta\phi(\text{jet}_2, E_T^{\text{miss}})\right)} \tag{6.2}$$

Tables 6.3 and 6.4 show the expected yields for all considered background sources and the benchmark signal points, normalised to 36.07 fb^{-1}. The significance values quoted in the tables are calculated assuming $N_{\text{exp}} = N_S + N_B$, where N_{exp} is the expected number of observed events, and with a background uncertainty assumed to be 30% for SRA and 25% for SRB as a rough estimate for optimisation purposes only. Background estimates are taken directly from MC, and systematic uncertainties are not included in the quoted uncertainties. For details on the full background estimation, including the normalisation and extrapolation from control and validation regions into the signal regions, see Section 6.5.

Signal Region C

SRC is designed to enhance sensitivity to $\tilde{t}_1 \to t\tilde{\chi}_1^0$ decays where the mass difference between the $\tilde{t}_1$ and $\tilde{\chi}_1^0$ is approximately equal to the top mass. This is a particularly challenging regime, as the stop pairs are mostly produced back-to-back, meaning the missing momentum from the two $\tilde{\chi}_1^0$s is equal

	SRA-TT	**SRA-TW**	**SRA-T0**
Z + jets	2.15 ± 0.27	4.20 ± 0.50	8.62 ± 0.65
Dibosons	0.00 ± 0.00	0.35 ± 0.25	0.00 ± 0.00
$t\bar{t}$	0.60 ± 0.10	0.45 ± 0.12	1.45 ± 0.31
Single top	1.03 ± 0.55	0.60 ± 0.15	2.52 ± 1.02
$t\bar{t} + Z$	2.46 ± 0.28	1.43 ± 0.21	2.02 ± 0.21
W + jets	0.65 ± 0.16	0.70 ± 0.20	1.58 ± 0.47
Total MC	6.88 ± 0.70	7.72 ± 0.65	16.20 ± 1.35
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^0})$ = (1000,1) GeV	8.21 ± 0.37 (1.8σ)	4.61 ± 0.27 (0.9σ)	6.43 ± 0.33 (0.7σ)

Table 6.3: Expected background and signal yields for SRA, for 36.07 fb^{-1} luminosity. Uncertainties are from MC statistics only, and the quoted significances (σ) assume 30% background uncertainty.

	SRB-TT	**SRB-TW**	**SRB-T0**
Z + jets	7.71 ± 0.66	14.36 ± 1.07	53.58 ± 1.79
Dibosons	0.13 ± 0.13	0.60 ± 0.35	1.03 ± 0.56
$t\bar{t}$	6.09 ± 0.59	12.80 ± 1.03	47.19 ± 2.04
Single top	3.59 ± 0.64	5.09 ± 0.46	22.43 ± 1.59
$t\bar{t} + Z$	7.25 ± 0.50	8.37 ± 0.51	15.92 ± 0.62
W + jets	6.12 ± 2.74	3.83 ± 0.62	20.40 ± 3.15
Total MC	30.89 ± 2.99	45.05 ± 1.78	160.57 ± 4.53
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^0})$ = (600,300) GeV	20.42 ± 1.05 (1.7σ)	25.96 ± 1.13 (1.6σ)	43.54 ± 1.39 (0.8σ)

Table 6.4: Expected background and signal yields for SRB, for 36.07 fb^{-1} luminosity. Uncertainties are from MC statistics only, and the quoted significances (σ) assume 25% background uncertainty.

and opposite. As a result, the E_T^{miss} is very small, and the signal looks indistinguishable from the non-resonant $t\bar{t}$ background. In order to combat this, a high-p_T ISR system is required in the final state, which recoils against the $\tilde{t}_1$ pair and boosts them in the transverse plane. This new event topology provides some unique kinematics that can be exploited to separate signal and background [97, 98].

A recursive jigsaw reconstruction technique (see Appendix A [99]) is used to divide each event into an ISR hemisphere and a sparticle hemisphere. The advantage of this method is that it does not place any requirement on the number of ISR jets - the alternative, searching for a single asymmetrically large-p_T jet to define as the ISR can be highly inefficient. By using the recursive jigsaw technique,

events with any number of arbirtrary ISR emissions can be described, improving the sensitivity to these compressed stop decays. Objects are grouped together based on their proximity in the lab frame's transverse plane. The two hemispheres are then chosen by minimising the transverse mass of all choices of object assignment simultaneously. Kinematic variables are defined based on this assignment of objects to either the ISR system or the sparticle system (denoted as S).

In addition to the preselection criteria, the requirements for SRC are shown in Table 6.5. Events are divided into 5 bins of the R_{ISR} variable, which is defined as:

$$R_{\text{ISR}} = \frac{E_T^{\text{miss}}}{p_T^{ISR}} \sim \frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{t}_1}} \tag{6.3}$$

In other words, R_{ISR} is the ratio of the missing energy to the transverse momentum of the ISR jet, which is proportional to the ratio of the two sparticle masses. By dividing the signal region into non-overlapping bins of R_{ISR}, each bin can target different signal models that peak in R_{ISR} within that bin range. The bin with the highest observed significance for each signal mass point is then used for the exclusion fit, to maximise sensitivity. The other variables used in this signal region are:

- N_{jet}^S: the number of jets in the sparticle hemisphere,
- N_{btag}^S: the number of b-tagged jets in the sparticle hemisphere,
- $p_{T,b}^{0,S}$: p_T of the leading b-tagged jet in the sparticle hemisphere,
- $p_T^{4,S}$: p_T of the fourth highest-p_T jet in the sparticle hemisphere,
- $\Delta\phi_{\text{ISR},I}$: angular separation in ϕ between the ISR system and the E_T^{miss} in the centre-of-mass frame,
- p_T^{ISR}: p_T of the ISR system, evaluated in the centre-of-mass frame,
- and m_T^S: transverse mass of the whole sparticle system plus the E_T^{miss}.

Table 6.6 shows the expected background and signal yields, taken from MC and normalised to 36.07 fb^{-1} luminosity. Again, no scale factors from the control or validation regions have been applied at this stage - Section 6.5 details this process including the definition of the control regions.

Variable	SRC-1	SRC-2	SRC-3	SRC-4	SRC-5
b-tagged jets	≥ 1				
N^S_{btag}	≥ 1				
N^S_{jet}	≥ 5				
$p^{0,S}_{T,b}$	> 40 GeV				
m^S_T	> 300 GeV				
$\Delta\phi_{\text{ISR},I}$	> 3.00 radians				
p^{ISR}_T	> 400 GeV				
$p^{4,S}_T$	> 50 GeV				
R_{ISR}	0.30-0.40	0.40-0.50	0.50-0.60	0.60-0.70	0.70-0.80

Table 6.5: Selection criteria for SRC in addition to the preselection requirements from Table 6.1. Events are divided into 5 bins of R_{ISR}, labelled SRC1-5.

Signal Region D

Signal region D is designed to target events where both top squarks decay via $\tilde{t}_1 \to b\tilde{\chi}^{\pm}_1 \to bW^{(*)}\tilde{\chi}^0_1$. The models considered during the optimisation of this SR have the chargino mass fixed to twice that of the neutralino, $m(\tilde{\chi}^{\pm}_1) = 2m(\tilde{\chi}^0_1)$, to satisfy gaugino universality. The event topology is similar to that of SRA and SRB, hence some of the same variables are used - however a much larger emphasis is on the jet p_T than in either of those signal regions. Two versions of the SR are defined, SRD-low and SRD-high, which are optimised separately for signals with smaller (SRD-low, ~ 300 GeV) or larger (SRD-high, ~ 500 GeV) values of the $\tilde{t}_1$-$\tilde{\chi}^0_1$ mass splitting. Table 6.7 describes the selection criteria for this signal region, and Table 6.8 shows the expected background and signal yields, taken from MC with no scale factors applied.

	SRC-1	**SRC-2**	**SRC-3**
Z + jets	0.46 ± 0.09	0.90 ± 0.13	0.74 ± 0.15
Dibosons	0.39 ± 0.28	0.21 ± 0.21	0.28 ± 0.28
$t\bar{t}$	18.19 ± 1.57	31.20 ± 1.82	20.62 ± 1.07
Single top	1.44 ± 0.51	1.02 ± 0.19	1.04 ± 0.41
$t\bar{t} + Z$	0.22 ± 0.10	0.46 ± 0.19	0.44 ± 0.11
W + jets	0.64 ± 0.13	1.53 ± 0.31	1.51 ± 0.37
Total MC	21.34 ± 1.69	35.30 ± 1.88	24.64 ± 1.25
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^0})$ = (300,127) GeV	30.68 ± 4.17 (5.5σ)	72.20 ± 7.29 (9.7σ)	14.80 ± 2.56 (2.7σ)
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^0})$ = (400,227) GeV	1.57 ± 0.45 (0.3σ)	10.81 ± 1.00 (1.7σ)	30.01 ± 1.67 (5.2σ)
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^0})$ = (500,327) GeV	0.11 ± 0.06 (0.0σ)	1.42 ± 0.26 (0.2σ)	6.90 ± 0.56 (1.3σ)

	SRC-4	**SRC-5**
Z + jets	0.45 ± 0.09	0.09 ± 0.03
Dibosons	0.00 ± 0.00	0.00 ± 0.00
$t\bar{t}$	6.95 ± 0.46	0.89 ± 0.24
Single top	0.62 ± 0.17	0.00 ± 0.25
$t\bar{t} + Z$	0.07 ± 0.08	0.05 ± 0.09
W + jets	1.53 ± 0.41	0.17 ± 0.09
Total MC	9.61 ± 0.65	1.12 ± 0.36
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^0})$ = (300,127) GeV	0.80 ± 0.57 (0.1σ)	0.55 ± 0.39 (0.2σ)
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^0})$ = (400,227) GeV	8.95 ± 0.86 (2.5σ)	0.43 ± 0.17 (0.1σ)
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^0})$ = (500,327) GeV	10.53 ± 0.68 (2.9σ)	1.19 ± 0.22 (0.8σ)

Table 6.6: Expected background and signal yields for SRC, for several signal mass points, for 36.07 fb^{-1} of data. Uncertainties are MC statistics only; for the purpose of the significance (σ) calculation, a 20% background uncertainty is assumed.

	SRD-low	SRD-high
E_T^{miss}	> 250 GeV	
N_{jets}	≥ 5	
b-tagged jets	≥ 2	
$\Delta R(b,b)$	> 0.8	
τ-veto	yes	
jet $p_T^{0,1}$	> 150 GeV	
jet p_T^3	> 100 GeV	> 80 GeV
jet p_T^4	> 60 GeV	
$m_T^{b,\text{min}}$	> 250 GeV	> 350 GeV
$m_T^{b,\text{max}}$	> 300 GeV	> 450 GeV
b-jet $p_T^1 + p_T^2$	> 300 GeV	> 400 GeV

Table 6.7: Selection requirements for SRD. SRD-low targets $\tilde{t}_1 \to b\tilde{\chi}_1^{\pm} \to bW^{(*)}\tilde{\chi}_1^0$ decays with small $\Delta\, m(\tilde{t}_1, \tilde{\chi}_1^0)$, and SRD-high targets decays with large $\Delta\, m(\tilde{t}_1, \tilde{\chi}_1^0)$.

	SRD-low	**SRD-high**
Z + jets	6.67 ± 0.44	3.10 ± 0.27
Dibosons	0.00 ± 0.00	0.00 ± 0.00
$t\bar{t}$	3.43 ± 0.37	1.04 ± 0.20
Single top	3.30 ± 0.47	1.30 ± 0.22
$t\bar{t} + Z$	3.06 ± 0.31	1.06 ± 0.15
W + jets	4.78 ± 2.68	0.84 ± 0.16
Total MC	21.25 ± 2.80	7.34 ± 0.46
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^{\pm}}, m_{\tilde{\chi}_1^0}) = (400,100,50)$ GeV	33.33 ± 6.07 (3.4σ)	-
$(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^{\pm}}, m_{\tilde{\chi}_1^0}) = (700,200,100)$ GeV	-	10.49 ± 1.15 (2.4σ)

Table 6.8: Expected background and signal yields for SRD-low and SRD-high for 36.07 fb^{-1}. No scale factors are applied, and uncertainties are from MC statistics only. The significances (σ) quoted assume a 25% background uncertainty.

6.5 Background Estimation

The background sources considered in this analysis were covered in Section 6.1. As the signal region data are blinded, control regions (CRs) are used instead to make sure that the background is well-understood, and to extract the background estimation for the SRs. Control regions are designed to enhance a particular source of background while being orthogonal to the SRs and containing as little signal as possible, demonstrated in Figure 6.8. The CRs are used to normalise the MC samples to data, but the shape of the distributions and the extrapolation of the background yields from the CRs to the SRs is taken from simulation. Sufficient data are needed to avoid large statistical uncertainties in the background estimates, and the CR definitions are chosen to be as kinematically close to the SRs as possible, to minimise the systematic uncertainties associated with the background yield extrapolation between the two regions. In some cases, the CR definition is too different from the SR to do the extrapolation without significant systematic uncertainty - in those cases, a validation region is defined in order to cross-check the extrapolation and to reduce the systematic uncertainty. As each CR is not purely the intended background process, the cross-contamination between CRs from other processes is estimated with simulated samples. All normalisation factors and the cross-contamination are determined simultaneously using a fit to each control, validation and signal region, as described below.

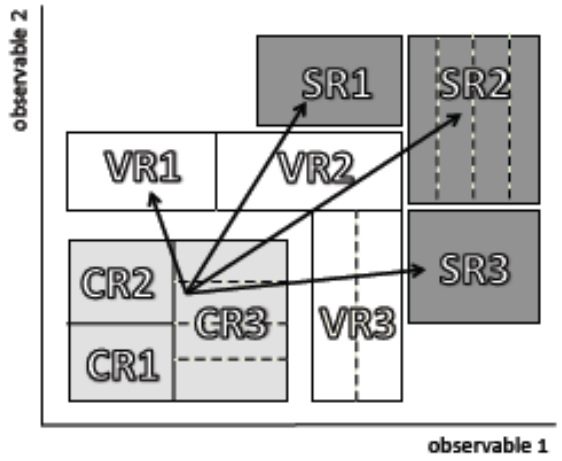

Figure 6.8: A diagrammatic representation of the separation between control, validation and signal regions, in some two-dimensional phase space. In the multi-dimensional phase space of this analysis, this sort of visualisation is much more difficult, but the general principle is the same - no control region should overlap with a signal region! [100]

Table 6.9 shows the preselection requirements common to all CRs. These differ from the preselection

criteria for the SRs, mostly because the CRs require the presence of at least one lepton. The exception to these preselection requirements are the photon control regions, for γ + jets and $t\bar{t}+\gamma$, which are used as an alternate method of estimating Z + jets and $t\bar{t}+Z$ backgrounds. In the 2-lepton CRs, the p_T of the leptons are added to the E_T^{miss}, treating the leptons as invisible to replicate $Z \to \nu\nu$ decays - this recalculated missing energy is called $E_T^{\text{miss}\prime}$.

	1 Lepton	2 Lepton
N_{lep}	1	2
N_{jets}	≥ 4	≥ 4
N_{btag}	≥ 1	≥ 1
$\lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert$	> 0.4	-
E_T^{miss}	> 250 GeV	< 50 GeV
$E_T^{\text{miss}\prime}$	-	> 100 GeV
$m_T(l, E_T^{\text{miss}})$	< 100 GeV	-
m_{ll}	-	> 86, < 96 GeV
$p_T^{\text{jet }0,1}$	> 80 GeV	> 40 GeV
$p_T^{\text{jet }2,3}$	> 40 GeV	> 40 GeV
Trigger	E_T^{miss}	Lepton

Table 6.9: Preselection requirements for the Control Regions. The trigger refers to either the same E_T^{miss} triggers used for the SRs, or to the lepton triggers described in Section 6.2.

The all-hadronic $t\bar{t}$ and QCD multijet backgrounds are found to be negligible for this analysis. No specific control region has been designed for them, but a data-driven method is used to estimate their contributions. The methodology for this is described in detail in Appendix B and Ref. [101], though the summary is that a jet response map is derived from simulated dijet events, then this response function is used to smear the jet-p_T in low-E_T^{miss} seed events to obtain pseudo-data that can be used for a yield estimation. The jet response is cross-checked with data where the E_T^{miss} can be unambiguously attributed to the mis-measurement of one of the jets. Diboson production and $t\bar{t}+W$ production (but not $t\bar{t}+Z$), which are sub-dominant background processes, are estimated directly from simulation.

Table 6.10 shows which control regions are used in the estimation of the background for each signal region, to summarise the individual contributions. The validation regions use the same CRs as the associated SRs, and are not listed as a result.

Control region	SRA-TT	SRA-TW	SRA-T0	SRB-TT	SRB-TW	SRB-T0	SRC	SRD
CRZAB-TT-TW	✓	✓	-	✓	✓	-	-	-
CRZAB-T0	-	-	✓	-	-	✓	-	-
CRZD	-	-	-	-	-	-	-	✓
CRTopA-TT	✓	-	-	-	-	-	-	-
CRTopA-TW	-	✓	-	-	-	-	-	-
CRTopA-T0	-	-	✓	-	-	-	-	-
CRTopB-TT	-	-	-	✓	-	-	-	-
CRTopB-TW	-	-	-	-	✓	-	-	-
CRTopB-T0	-	-	-	-	-	✓	-	-
CRTopC	-	-		-	-	-	✓	-
CRTopD	-	-	-	-	-	-	-	✓
CRST	✓	✓	✓	✓	✓	✓	✓	✓
CRW	✓	✓	✓	✓	✓	✓	✓	✓
CRTTGamma	✓	✓	✓	✓	✓	✓	✓	✓

Table 6.10: Summary of which control region is used in the background estimation fitting procedure for which signal region. Where not otherwise labelled, each sub-region (e.g. SRD-low and SRD-high) uses the same control region.

Simultaneous Fit to Determine SM Background

The observed number of events in each of the various control regions are included in a profile likelihood fit [72] to determine the SM background estimates in each signal region. The likelihood function is built as the product of Poisson probability density functions, one for each control region, describing the expected and observed yields [100]. This procedure takes common systematic uncertainties, discussed in Section 6.6, between the control and signal regions and their correlations into account; these uncertainties are treated as nuisance parameters in the fit and are modelled by Gaussian probability density functions. Background processes that do not have dedicated CRs have their contributions fixed to the values expected by simulation, with their uncertainties used as nuisance parameters in the fit. Control region normalisation is performed first, then the normalised CR yield is used to extrapolate the expected background yield from the CR to SR using so-called 'transfer factors', T_f, the ratio of expected event yields between each region. The background estimation for each background source in each signal region is calculated as follows:

$$N_p(\text{SR,est.}) = N_p(\text{CR,obs.}) \times \left[\frac{\text{MC}_p(\text{SR,raw})}{\text{MC}_p(\text{CR,raw})}\right] = N_p(\text{CR,obs.}) \times T_f, \tag{6.4}$$

where N_p(SR,est.) is the SR estimation for a background process p, N_p(CR,obs.) is the number of observed events in the CR for the process p, and MC$_p$(SR(CR),raw) are the raw MC estimates in the SR (CR), taken pre-fit.

The background estimates are validated by using the simultaneous fit to estimate the background in dedicated validation regions and comparing to observation. Validation regions (VRs) are designed to be orthogonal to both the CR and SR while retaining kinematics and event composition close to the SRs, but with little contribution from any of the signal models considered. Normalisation factors from the VRs are used in the fit to correct the transfer factor from CR to SR.

Ideally, a specific control region would be designed for each background process for each signal region, so that the background contribution in each SR can be estimated independently of the other SRs. This has been done for the Z + jets and $t\bar{t}$ background sources, but due to limited statistics, only a single CR exists for the W + jets, single top and $t\bar{t}$ + Z processes. For these three CRs, a choice has to be made as to which of the CRZ and CRTop regions should be used in the fit, to constrain the normalisation of the Z and $t\bar{t}$ backgrounds in these CRs. In CRW, the top candidate mass criterion rules out any CRZ or CRTop region whose definition involves an ISR jet or a top mass requirement, leaving only CRZD and CRTopD to be included in the fit. In CRST, it was found that the majority of events fell into the T0 top category, resulting in the choice of CRZAB-T0 for the Z normalisation and CRTopB-T0 for the $t\bar{t}$ normalisation. Finally in CRTTGamma, CRZAB-T0 and CRTopB-T0 are chosen again, for the same reasons as for CRST. To reiterate: these choices are made only during the process of normalising the CRW, CRST and CRTTGamma control regions - all other regions are included in the rest of the fitting procedure.

Z + Jets Control Regions

Three CRs are defined for the purpose of estimating $Z \to \nu\nu$ + jets. SRA and SRB each share two control regions, CRZAB-TT-TW and CRZAB-T0, which are designed to estimate the background separately in the T0 and TT+TW top reconstruction categories. SRD has a single control region, CRZD, to maximise the statistics in that region. No control region for SRC has been defined, due to the very small contribution from Z+jets in that region - instead the background estimation is taken from simulation. Table 6.11 shows the selection criteria for each Z CR. Figures 6.9 and 6.10 show distributions for some of the discriminating variables in these CRs - agreement between data and MC is good in all variables across all regions.

Selection	CRZAB-TT-TW	CRZAB-T0	CRZD
Preselection	2 Leptons		
Jet multiplicity	≥ 4		≥ 5
Jet p_T	(80,80,40,40) GeV		
N_{btag}	≥ 2		
$m^0_{\text{jet},R=1.2}$	> 120 GeV		-
$m^1_{\text{jet},R=1.2}$	> 60 GeV	< 60 GeV	-
$m_T^{b,\text{min}\prime}$	-		200
$m_T^{b,\text{max}\prime}$	-		200

Table 6.11: Selection criteria for all Z CRs. Any variable name ending with $\prime$ indicates the variable is calculated with the lepton p_T added to the E_T^{miss}.

Z + Jets Validation Regions

As the presence of leptons in the control regions makes them too different from the signal regions to be able to extrapolate without significant uncertainty, validation regions are defined. The VRs, VRZAB for SRA and SRB, and VRZD for SRD, are all-hadronic, vetoing any events containing leptons. The variables used are the same as in the SRs, but with looser requirements - to maintain orthogonality with the SRs, the $\Delta R(b,b)$ and/or the $m^0_{\text{jet},R=1.2}$ requirements are reversed. This reversal also helps to reduce signal contamination, which is below 25% for all VRs, with the signal causing the highest contamination having $m_{\tilde{t}_1} = 500$ GeV and $m_{\tilde{\chi}^0_1} = 200$ GeV (an already excluded signal point). The majority of signal points, including all non-excluded signal points, are below 15% contamination in all validation regions. Table 6.12 contains the selection criteria for the Z VRs - the same preselection as for the SRs is applied in addition to these requirements. Figures 6.11 and 6.12 show the distributions of various discriminating variables in the three regions. After the simultaneous fit, the agreement between data and MC is good across all regions and variables.

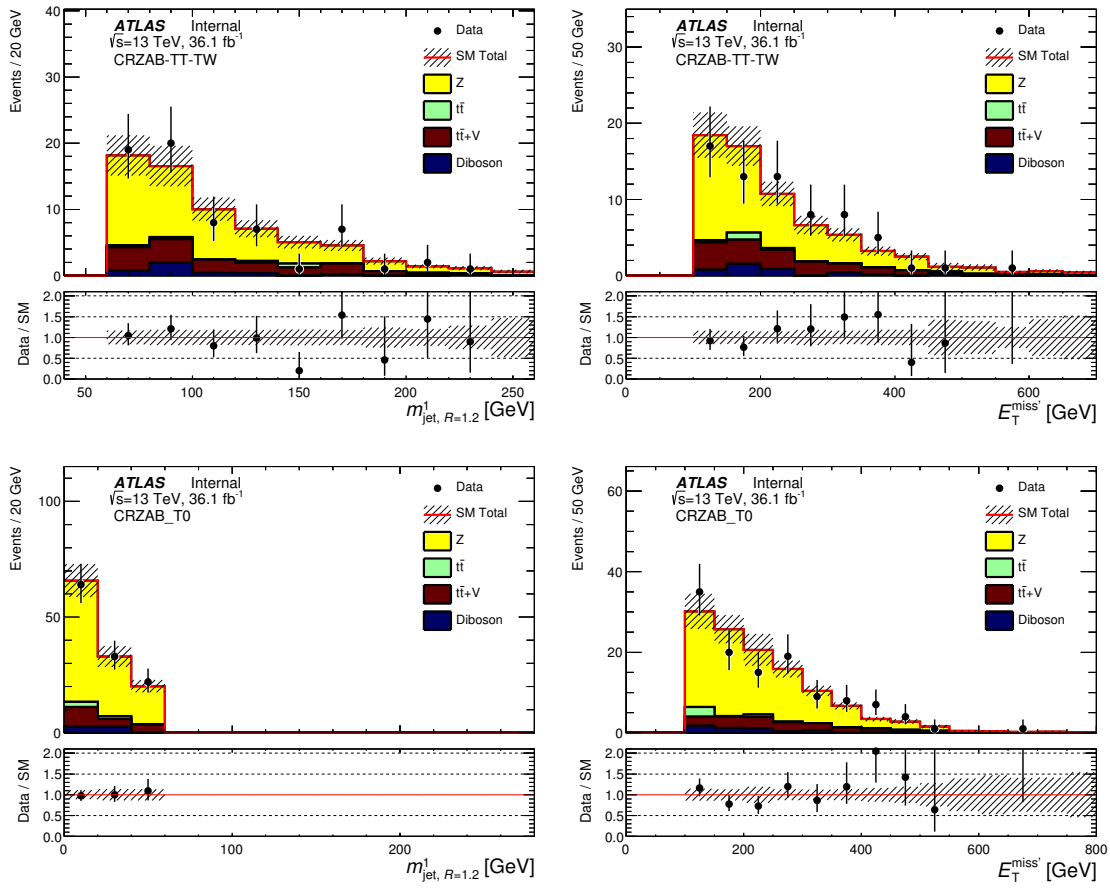

Figure 6.9: Distributions of $m^1_{\text{jet},R=1.2}$ and $E_T^{\text{miss}\prime}$ for CRZAB-TT-TW and CRZAB-T0, after the simultaneous fit across CRs has normalised the simulation to data. The hashed area represents the uncertainty due to MC statistics and detector systematics.

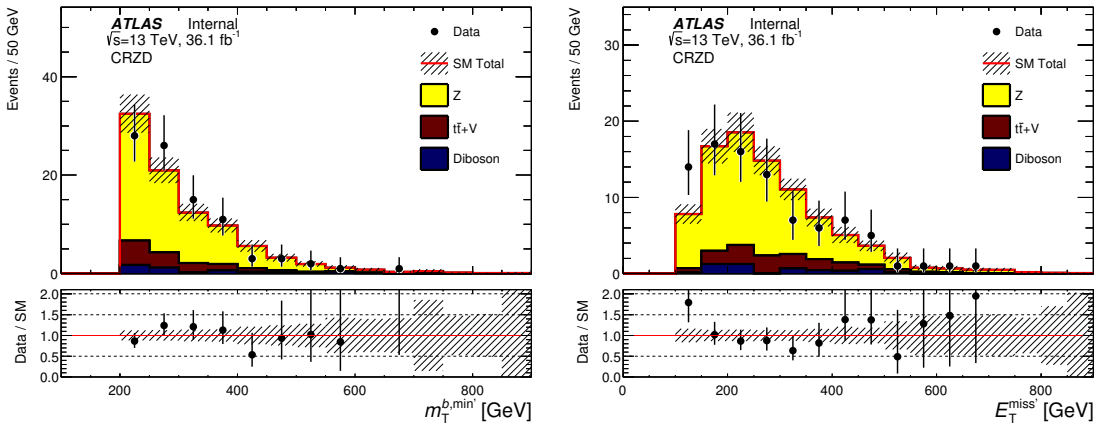

Figure 6.10: Distributions of $m_T^{b,\min\prime}$ and $E_T^{\text{miss}\prime}$ for CRZD, after the simultaneous fit across CRs has normalised the simulation to data. The hashed area represents the uncertainty due to MC statistics and detector systematics.

Selection	VRZAB	VRZD
Jet p_T^0, p_T^1	80, 80 GeV	150, 80 GeV
Jet multiplicity	≥ 4	≥ 5
N_{btag}	>= 2	
τ veto	yes	
$m_T^{b,\min}$	> 200 GeV	
$m^0_{\text{jet},R=1.2}$	< 120 GeV	-
$\Delta R(b,b)$	< 1	< 0.8
$m_T^{b,\max}$	-	> 200 GeV

Table 6.12: Selection criteria for the Z validation regions. The same preselection as for the SRs is applied in addition to these requirements.

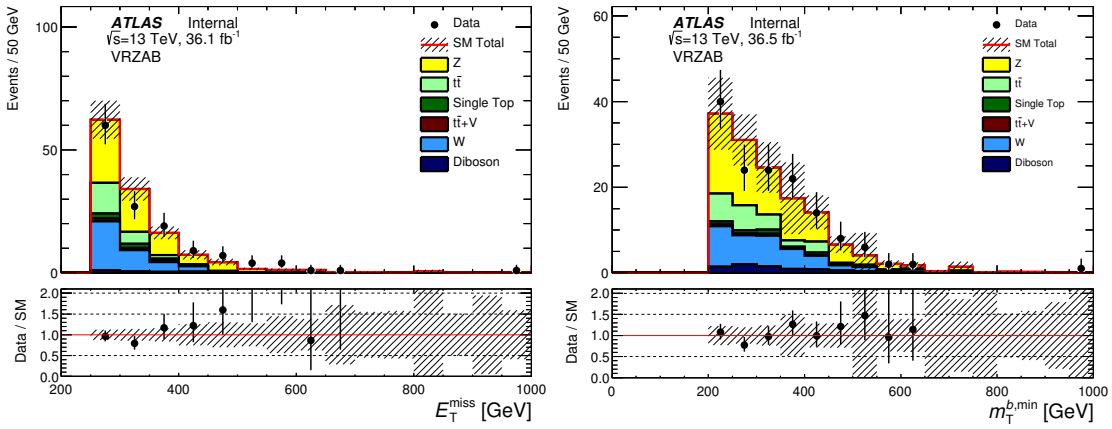

Figure 6.11: Distributions of E_T^{miss} and $m_T^{b,\text{min}}$ for VRZAB, for 36.07 fb^{-1} of data, after the simultaneous fit has been done to normalise simulation to data.

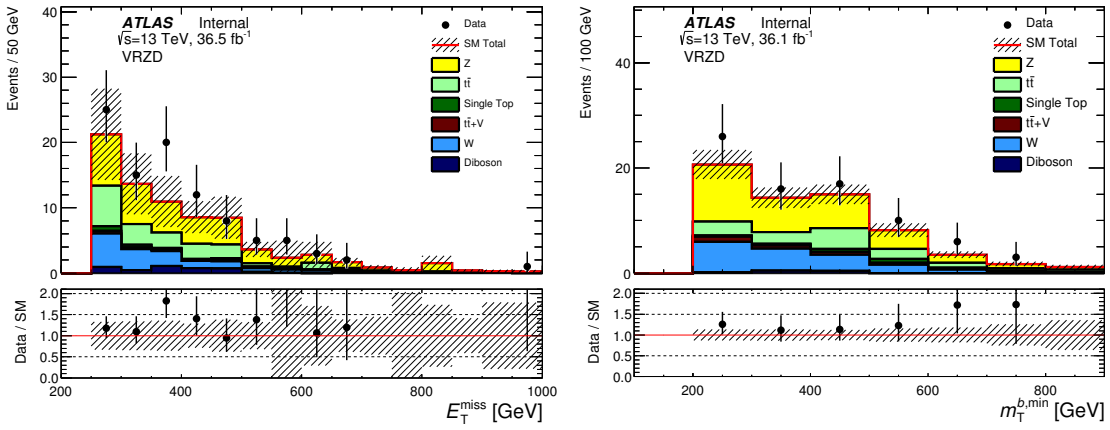

Figure 6.12: Distributions of E_T^{miss} and $m_T^{b,\text{min}}$ for VRZD, for 36.07 fb^{-1} of data, after the simultaneous fit has been done to normalise simulation to data.

One-lepton Control Regions

To estimate the W + jets, $t\bar{t}$ and single-top backgrounds, one-lepton control regions are defined. These backgrounds all contribute to the signal regions because one lepton from the decay of a W is either out of acceptance, mis-identified as a jet, or a hadronically-decaying τ. The CRs detailed in this section are used to estimate the normalisation of these backgrounds. For consistency with the signal regions, the same E_T^{miss} triggers are used as in the SR preselection shown in Table 6.1. The single-top and W + jets backgrounds have a single control region each, CRST and CRW respectively, while the $t\bar{t}$ background has several control regions - one each for SRC and SRD, and one for every SRA and SRB top reconstruction category. In each CR, the lepton is counted as a non-b-tagged jet for the p_T requirements and the jet reclustering, to emulate the hadronic τ decays, but not for the QCD cleaning requirements on $|\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})|$ and $m_T(l, E_T^{\text{miss}})$. The top control region for SRC, CRTopC, uses similar ISR and recursive jigsaw techniques as the signal region - for the definitions of those variables and how they were reconstructed, see the definition of SRC.

Table 6.13 contains the selection criteria for CRST and CRW, while Tables 6.14 and 6.15 contain the selection criteria for the CRTop regions. Distributions for a selection of discriminating variables are shown in Figures 6.13 through 6.18 for 36.07 fb^{-1} of data.

CRST, CRW and CRTop are mutually exclusive. The requirements on the number of b-jets and $m^0_{\text{jet},R=1.2}$ ensure that CRW is orthogonal to CRST and CRTop, while the selection on $\Delta R(b_{0,1}, l)_{\min}$, defined as the minimum ΔR between the two jets with the largest b-tag weight and the selected lepton, ensures the orthogonality of CRTop and CRST.

	CRST	**CRW**
Preselection	1 Lepton	
$m_T(l, E_T^{\text{miss}})$	> 30	
Number of b-jets	≥ 2	$= 1$
$m^0_{\text{jet},R=1.2}$	> 120 GeV	< 60 GeV
$m_T^{b,\min}$	> 200 GeV	-
$\Delta R(b_{0,1}, l)_{\min}$	> 2.0	
$\Delta R(b, b)$	> 1.5	-

Table 6.13: Summary of the selection for the 1-lepton single top and W+jets control regions. The lepton is treated as a jet for the jet counting and p_T ordering as well as for the top reconstruction. The 1-lepton preselection from Table 6.9 is applied.

While the signal contamination in CRW and all CRTop regions is negligible (less than 10%), in the

single-top control region the signal contamination is observed to be almost 25% for $\tilde{t}_1 \to b\tilde{\chi}_1^{\pm} \to bW^{(*)}\tilde{\chi}_1^0$ models.

CRTop		TT	TW	T0
	$m^0_{\text{jet},R=1.2}$	> 120 GeV		
	$m^1_{\text{jet},R=1.2}$	> 120 GeV	$60 - 120$ GeV	< 60 GeV
	$m_T(l, E_T^{\text{miss}})$	$> 30, < 100$ GeV		
	Number of b-jets	≥ 2		
	$m_T^{b,\text{min}}$	> 100 GeV		
	$\Delta R(b_{0,1}, l)_{\text{min}}$	< 1.5		
A	$m^0_{\text{jet},R=0.8}$	> 60 GeV		
	$\Delta R(b, b)$	> 1	-	
	E_T^{miss}	> 250 GeV	> 300 GeV	> 350 GeV
B	$m_T^{b,\text{max}}$	> 200 GeV		
	$\Delta R(b, b)$	> 1.2		

Table 6.14: Selection criteria for CRTopA-TT, CRTopA-TW, CRTopA-T0, CRTopB-TT, CRTopB-TW and CRTopB-T0. The 1-lepton preselection from Table 6.9 is applied in addition.

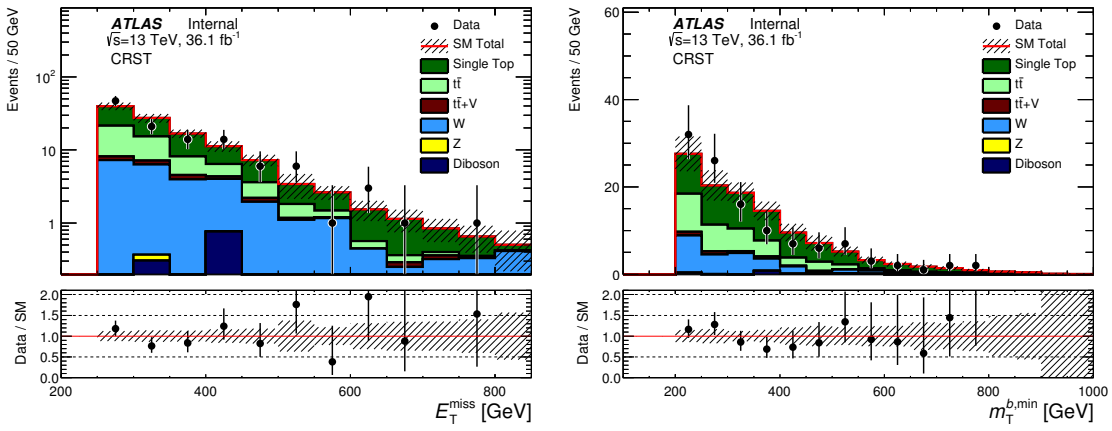

Figure 6.13: Distributions of E_T^{miss} and $m_T^{b,\text{min}}$ in CRST, for 36.07 fb^{-1} of data, after the simultaneous fit across all CRs to normalise simulation to data. The hashed area is the uncertainty due to MC statistics and detector systematics.

	CRTopC	**CRTopD**
$m_T(l, E_T^{\text{miss}})$	< 80 GeV	$> 30, < 100$ GeV
Number of b-jets	≥ 1	≥ 2
$m_T^{b,\text{min}}$	-	> 100 GeV
$\Delta R(b_{0,1}, l)_{\text{min}}$	< 2.0	< 1.5
N_{jet}^S	≥ 5	-
$N_{\text{b-tag}}^S$	≥ 1	-
$p_T^{4,S}$	> 40 GeV	-
p_T^{ISR}	≥ 400 GeV	-
$\Delta R(b, b)$	-	> 0.8
$m_T^{b,\text{max}}$	-	> 200 GeV
Jet $p_T^{0,1}$	-	> 150 GeV
Jet $p_T^{2,3}$	-	> 80 GeV
b-jet p_T^0+p_T^1	-	> 300 GeV

Table 6.15: Selection criteria for CRTopC and CRTopD. The 1-lepton preselection from Table 6.9 is applied in addition.

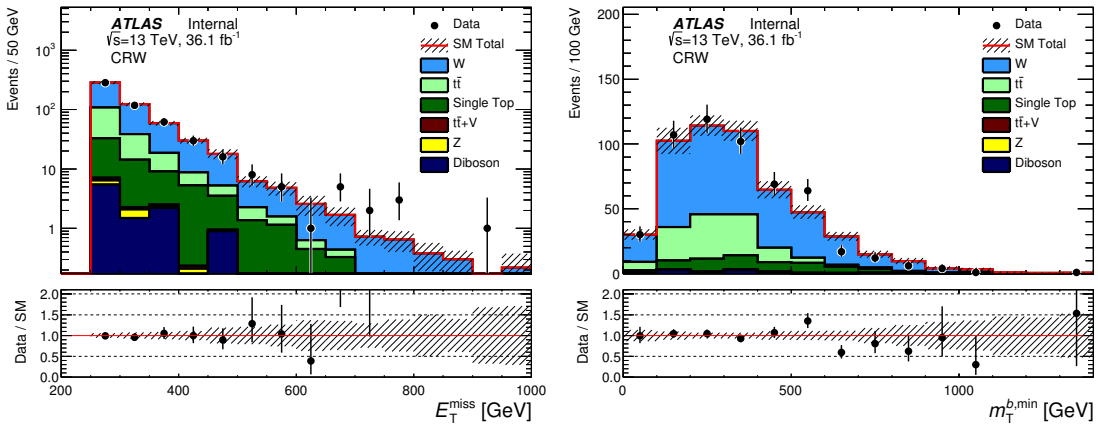

Figure 6.14: Distributions of E_T^{miss} and $m_T^{b,\text{min}}$ in CRW, for 36.07 fb^{-1} of data, after the simultaneous fit across all CRs to normalise simulation to data. The hashed area is the uncertainty due to MC statistics and detector systematics.

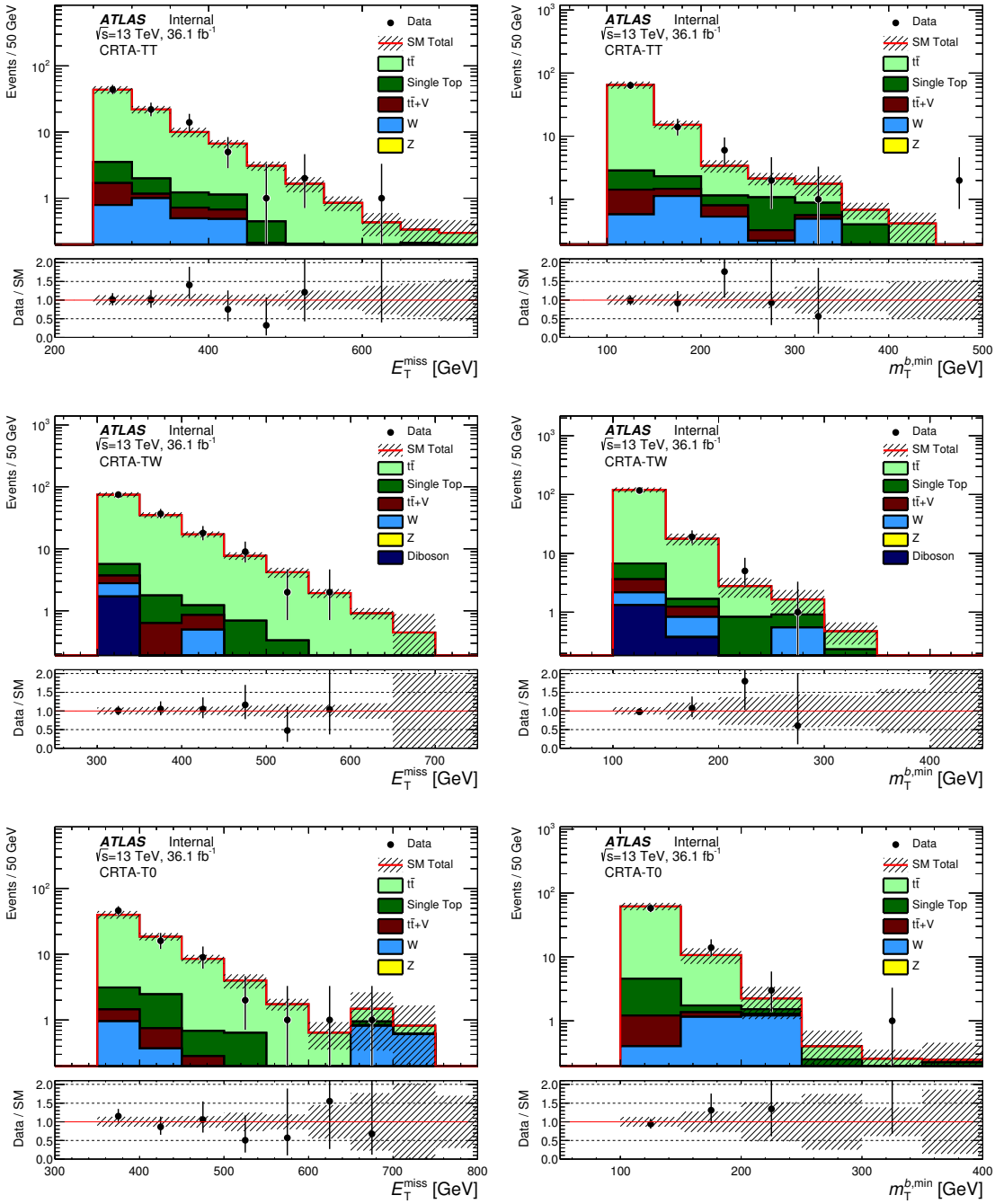

Figure 6.15: Distributions of E_T^{miss} and $m_T^{b,\text{min}}$ in CRTopA-TT, -TW and -T0, for 36.07 fb^{-1} of data, after the simultaneous fit across all CRs to normalise simulation to data. The hashed area is the uncertainty due to MC statistics and detector systematics.

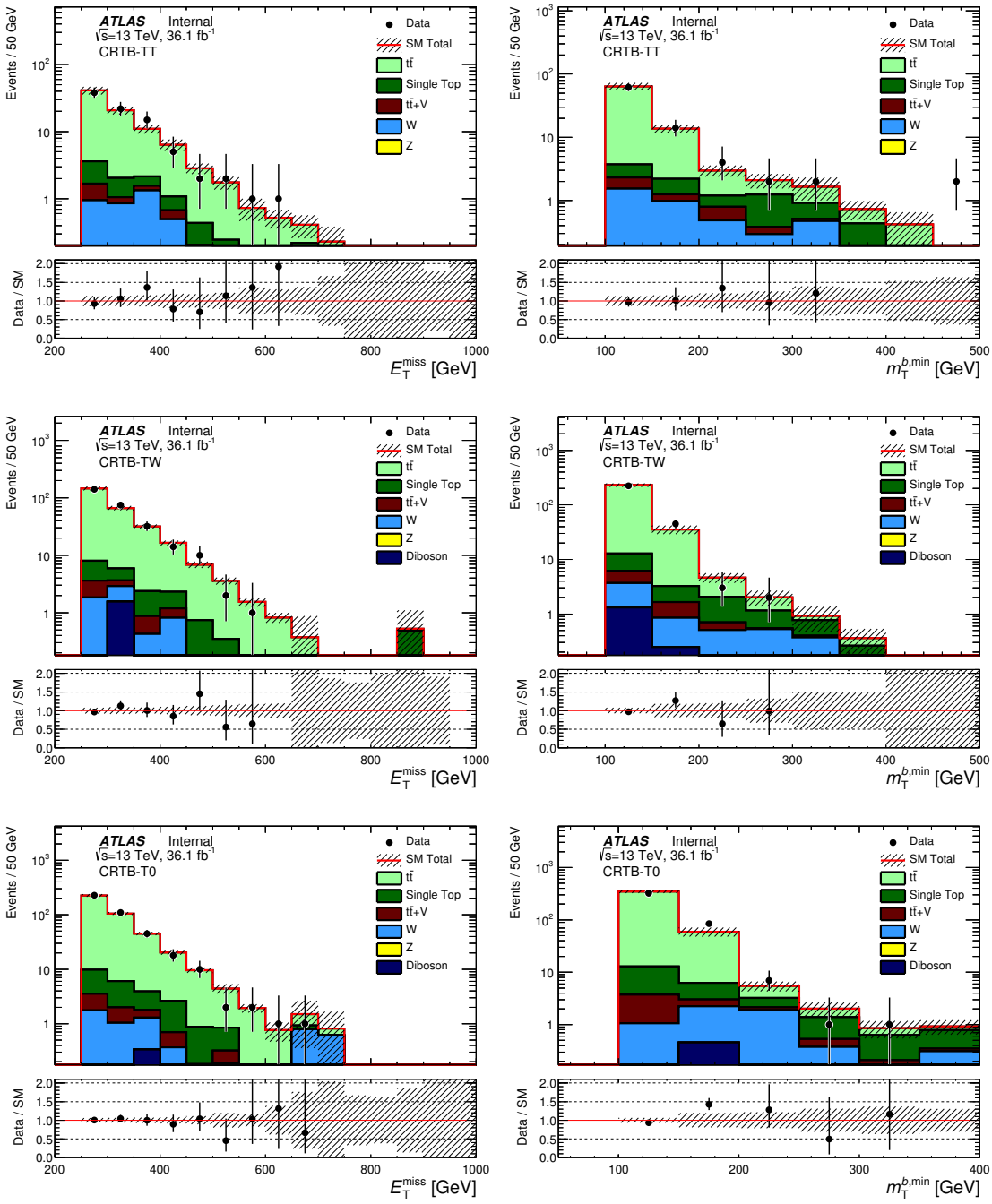

Figure 6.16: Distributions of E_T^{miss} and $m_T^{b,\text{min}}$ in CRTopB-TT, -TW and -T0, for 36.07 fb^{-1} of data, after the simultaneous fit across all CRs to normalise simulation to data. The hashed area is the uncertainty due to MC statistics and detector systematics.

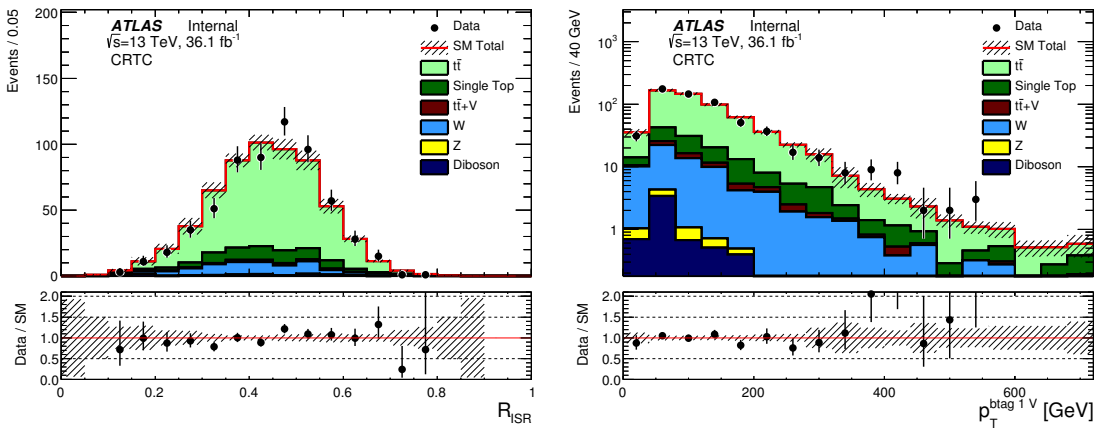

Figure 6.17: Distributions of R_{ISR} and $p_{T,b}^{0,S}$ in CRTopC, for 36.07 fb^{-1} of data, after the simultaneous fit across all CRs to normalise simulation to data. The hashed area is the uncertainty due to MC statistics and detector systematics.

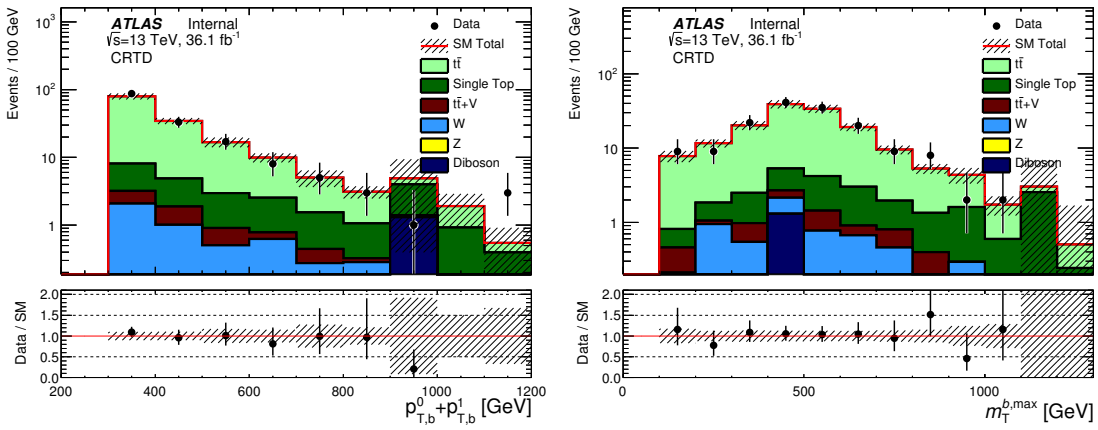

Figure 6.18: Distributions of b-jet $p_T^0 + p_T^1$ and $m_T^{b,\mathrm{max}}$ in CRTopD for 36.07 fb^{-1} of data, after the simultaneous fit to each control and validation region has normalised simulation to data. The hashed area shows the MC statistical and detector systematic uncertainties.

1-Lepton Validation Regions

Validation regions are required for the $t\bar{t}$ and W + jets backgrounds - the single top control region CRST is similar enough to the signal regions in terms to the topology being used, and the background a small enough contribution, that the estimation and normalisation can be taken from the control region directly. In all top validation regions, the same preselection as for the signal regions is applied, described in Table 6.1, however in the W validation region the 1-lepton preselection is applied, as described in Table 6.9. The top and W validation regions are defined in Tables 6.16 and 6.17. The purity of the $t\bar{t}$ VRs is $\sim$ 80%, with single-top and Z + jets processes contributing $\sim$ 7% each, and the purity of the W VR is 40% with single-top and $t\bar{t}$ processes contributing 30% each - signal contamination for both VRTop and VRW is 10% or less for all signal points. Figures 6.19 to 6.23 show distributions for some of the discriminating variables in the validation regions, for 36.07 fb^{-1} data - agreement between data and MC is good in all regions.

VRTop		**TT**	**TW**	**T0**
	Preselection	Signal		
	$m^{0}_{\text{jet},R=1.2}$	> 120 GeV		
	$m^{1}_{\text{jet},R=1.2}$	> 120 GeV	$60 - 120$ GeV	< 60 GeV
	$m_T^{b,\text{min}}$	$> 100, < 200$ GeV	$> 140, < 200$ GeV	$> 160, < 200$ GeV
	$N_{\text{b-tag}}$	≥ 2		
A	$m^{0}_{\text{jet},R=0.8}$	> 60 GeV		
	$\Delta R(b,b)$	> 1	-	
	E_T^{miss}	> 300 GeV	> 400 GeV	> 450 GeV
B	$\Delta R(b,b)$	> 1.2		
	$m_T^{b,\text{max}}$	> 200 GeV		

Table 6.16: Selection criteria for the VRATopA and VRATopB validation regions, for each top reconstruction category. The SR preselection described in Table 6.1 is also applied.

	VRTopC	**VRTopD**	**VRW**
Preselection	Signal		1 Lepton
$m_T^{b,\min}$	-	$> 100, < 200$ GeV	> 150 GeV
Jet multiplicity	≥ 4	≥ 5	≥ 4
$N_{\text{b-tag}}$	≥ 1	≥ 2	
N_{jet}^S	≥ 4	-	
$N_{\text{b-tag}}^S$	≥ 1	-	
$p_{T,b}^{0,S}$	≥ 40 GeV	-	
$p_T^{4,S}$	> 40 GeV	-	
p_T^{ISR}	≥ 400 GeV	-	
m_T^S	> 100 GeV	-	
m_V/m_S	< 0.6	-	
$\Delta\phi_{\text{ISR},I}$	< 3.00	-	
$\Delta R(b,b)$	-	> 0.8	-
$m_T^{b,\max}$	-	> 300 GeV	-
Jet $p_T^{0,1}$	-	> 150 GeV	-
Jet $p_T^{2,3}$	-	> 100 GeV	-
b-jet p_T^0+p_T^1	-	> 300 GeV	-
τ-veto	-	yes	-
$\lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert$	-		≥ 2
$m_T(l, E_T^{\text{miss}})$	-		$> 30, < 100$ GeV
$m_{\text{jet},R=1.2}^0$	-		< 70 GeV
$\Delta R(b_{0,1}, l)_{\min}$	-		> 1.8

Table 6.17: Selection criteria for VRTopC, VRTopD and VRW. The variable m_V/m_S is not used in SRC, but is defined as the ratio of the transverse mass of the visible part of the sparticle system to the transverse mass of the sparticle system plus the E_T^{miss}. For more details on the definition of the ISR and sparticle systems, see Appendix A.

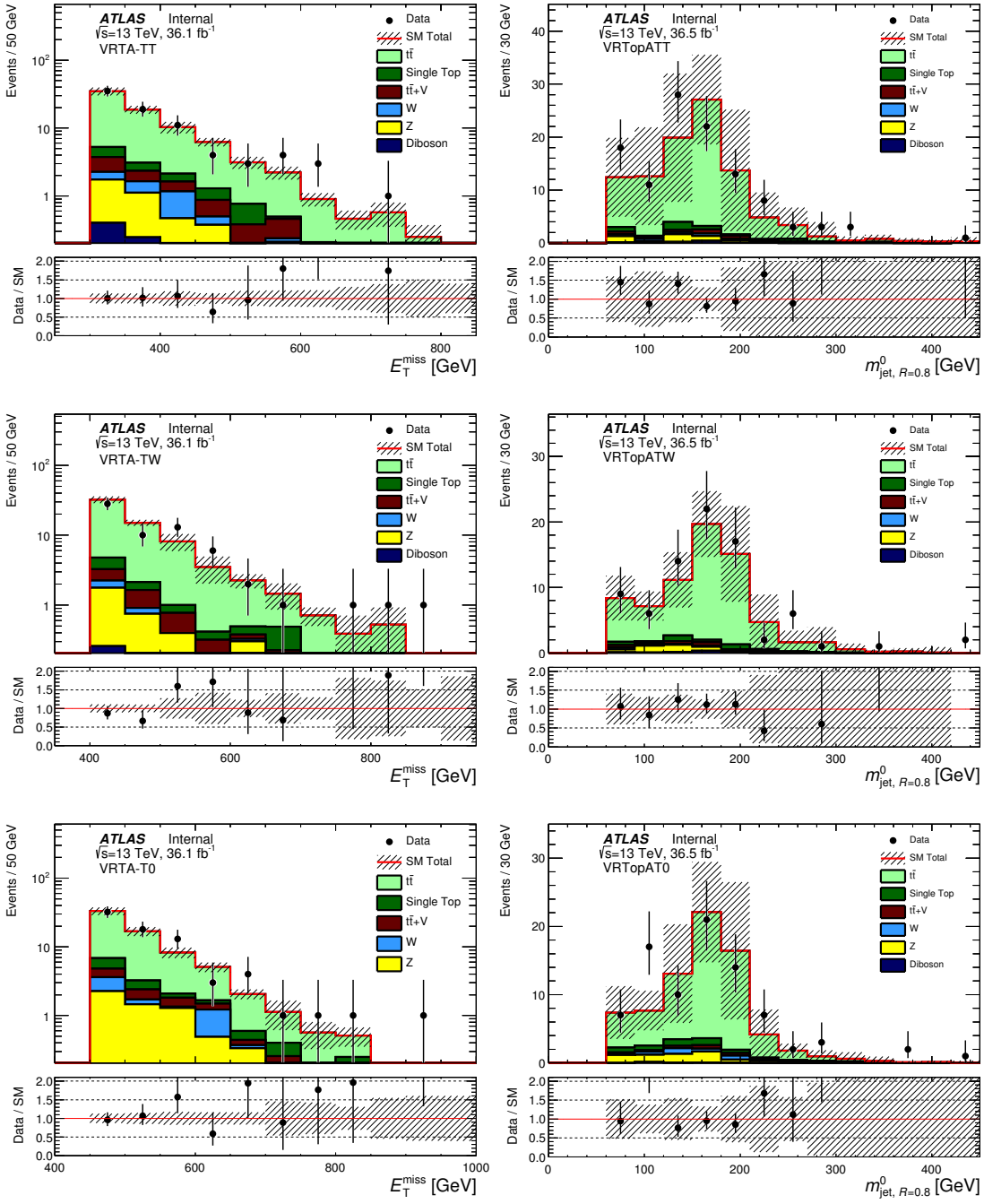

Figure 6.19: Distributions of E_T^{miss} and $m^0_{\text{jet},R=0.8}$ for VRTopA-TT, -TW and -T0 for 36.07 fb^{-1} of data, after the simultaneous fit to all control and validation regions has normalised simulation to data. The hashed area shows the MC statistical and detector systematic uncertainties.

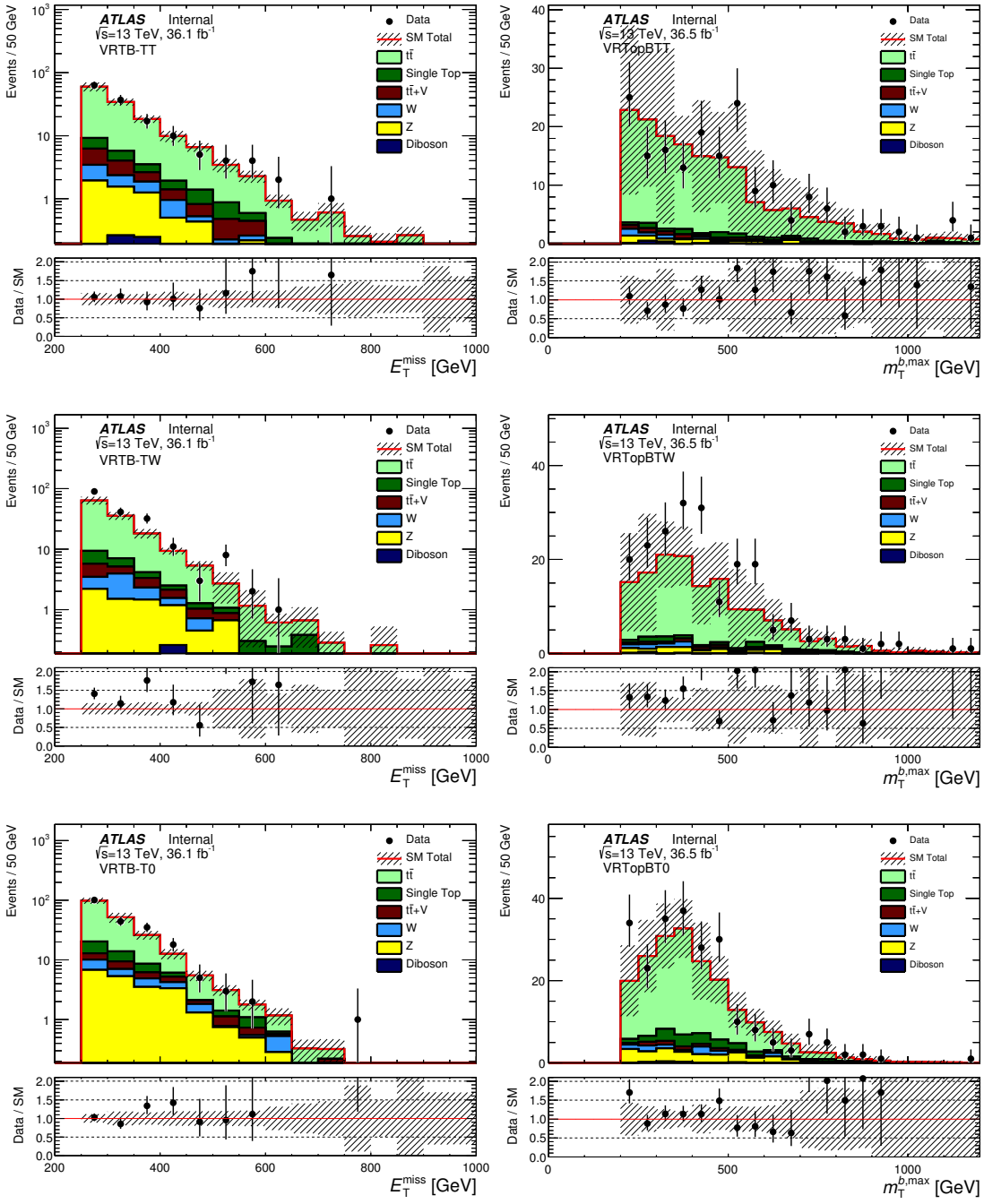

Figure 6.20: Distributions of E_T^{miss} and $m_T^{b,\text{max}}$ for VRTopB-TT, -TW and -T0 for 36.07 fb^{-1} of data, after the simultaneous fit to all control and validation regions has normalised simulation to data. The hashed area shows the MC statistical and detector systematic uncertainties.

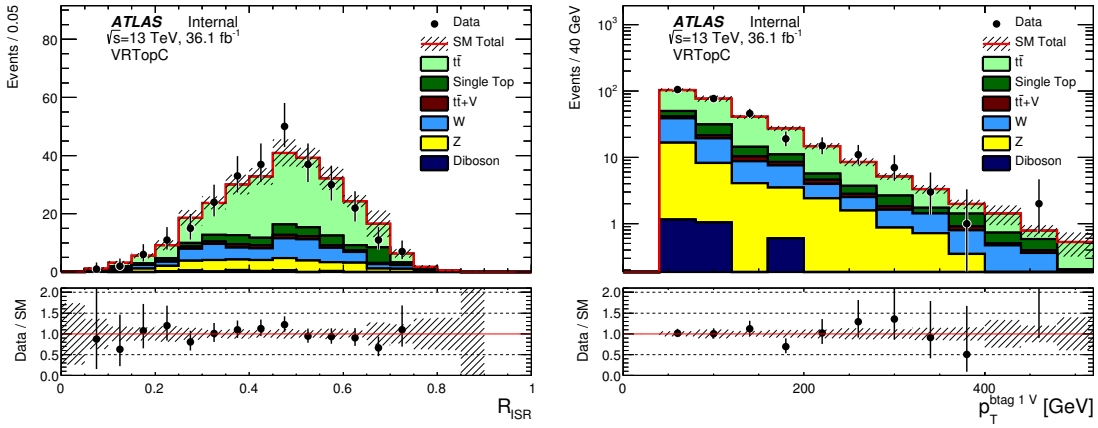

Figure 6.21: Distributions of R_{ISR} and $p_{T,b}^{0,S}$ in VRTopC for 36.07 fb^{-1} of data, after the simultaneous fit to all control and validation regions has normalised simulation to data. The hashed area shows the MC statistical and detector systematic uncertainties.

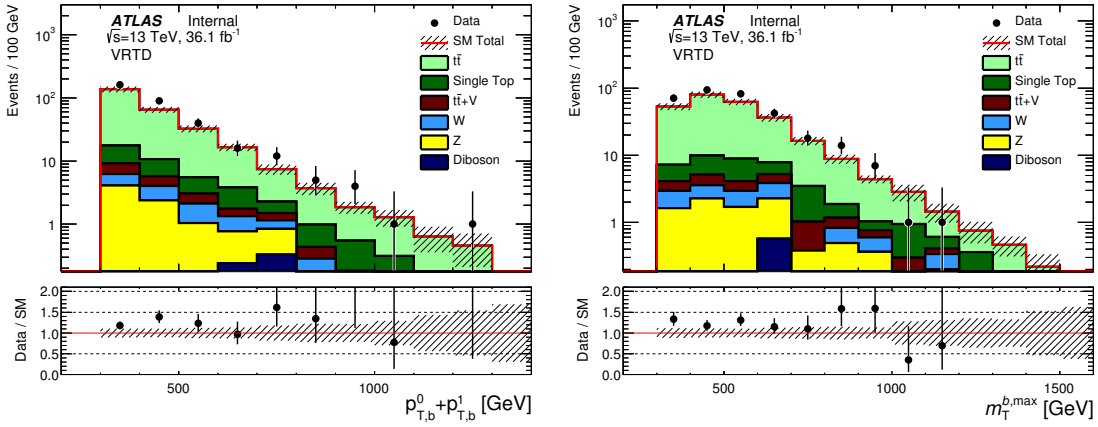

Figure 6.22: Distributions of b-jet $p_T^0 + p_T^1$ and $m_T^{b,\mathrm{max}}$ in VRTopD for 36.07 fb^{-1} of data, after the simultaneous fit to all control and validation regions has normalised simulation to data. The hashed area shows the MC statistical and detector systematic uncertainties.

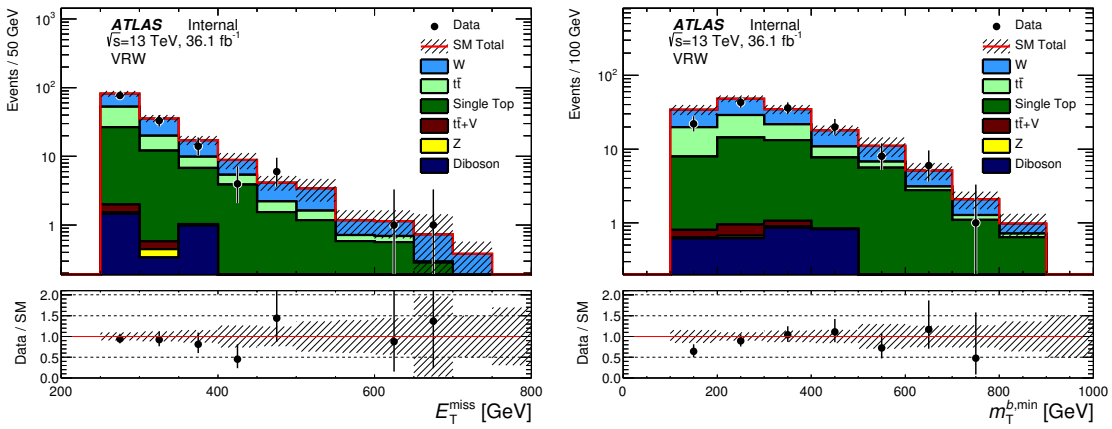

Figure 6.23: Distributions of E_T^{miss} and $m_T^{b,\text{min}}$ in VRW for 36.07 fb^{-1} of data, after the simultaneous fit to all control and validation regions has normalised simulation to data. The hashed area shows the MC statistical and detector systematic uncertainties.

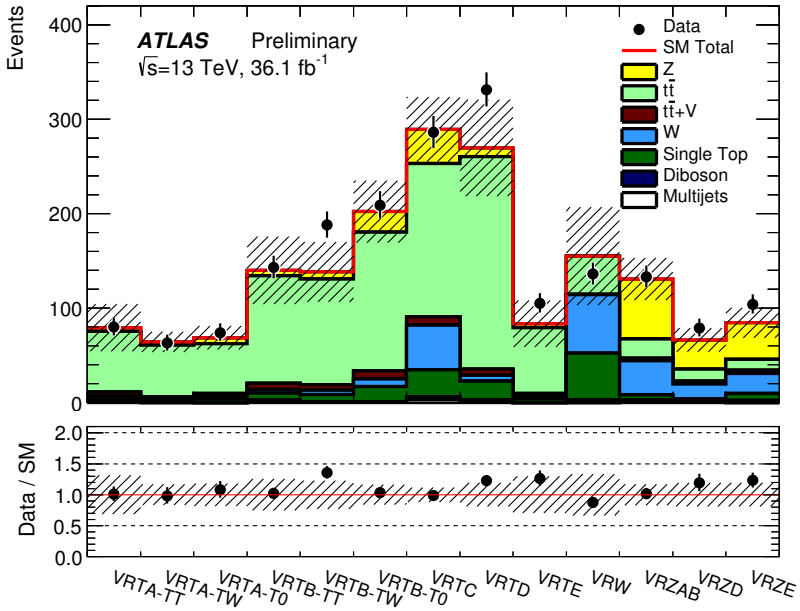

Figure 6.24: Final yields for all validation regions after the simultaneous fit to all control and validation regions. The stacked histograms show the SM expectation, with the hatched uncertainty band showing the total uncertainty including MC statistical uncertainty, detector-related systematics and theoretical uncertainties on the extrapolation from CR to VR. Both VRTE and VRZE are included here - these regions are covered in Chapter 7 as they are related to a reinterpretation of the hadronic stop data used in this chapter.

Estimating $t\bar{t} + Z$ using $t\bar{t} + \gamma$

The $t\bar{t} + Z(\rightarrow \nu\nu)$ background is an irreducible background that was small during the LHC Run 1, but has increased in $\sqrt{s} = 13$ TeV data. Designing a control region to target these decays explicitly, using charged lepton Z decays similarly to the Z + jets control regions, would be ideal. However such a CR is difficult to design due to low statistics and the small branching fraction ($\sim 3\%$ [3]) for $Z \rightarrow ll$ decays. In addition, any attempt to make a 2-lepton control region would suffer from a large contamination of $t\bar{t}$ and Z + jets processes. Instead, a different approach is taken, by creating a one-lepton CR for $t\bar{t} + \gamma$, which is a similar process. The γ is then used as a proxy for the invisibly-decaying Z boson, similarly to how in the Z control regions, the visible leptons are used to model the invisible neutrinos.

Table 6.18 describes the requirements for this control region, CRTTGamma. A zero-lepton validation region was considered for this CR as well, but it suffered from low statistics, and the 1-lepton CR used was found to be enough to minimise theoretical uncertainties in the extrapolation from the γ to the Z. The same lepton trigger requirements for the other 1-lepton CRs are used here, with the reconstructed lepton required to be matched to a trigger object and to have $p_T > 2$ GeV above the trigger lepton p_T threshold. Unlike in the other 1-lepton control regions, the lepton is not treated as a jet in order to model hadronic τ decays, nor is the lepton p_T added to the E_T^{miss} - instead the γ p_T is used to model the $E_T^{\text{miss}}$from the invisible Z decays. Figure 6.25 shows some distributions in this control region, and good data-MC agreement is observed.

Selection	CRTTGamma
Trigger	Single lepton
Leptons	exactly 1
Photons	exactly 1
Jet multiplicity	≥ 4
Jet p_T	(80,80,40,40) GeV
b-jet multiplicity	≥ 2
γ p_T	> 150 GeV

Table 6.18: Selection criteria for the $t\bar{t} + \gamma$ 1 lepton CR.

Estimating Z + Jets using γ + Jets

The method described previously for estimating the Z + jets background, by using the charged lepton Z decays to emulate the neutral lepton Z decays, is effective. Being able to look directly at $Z \rightarrow \nu\nu$ decays

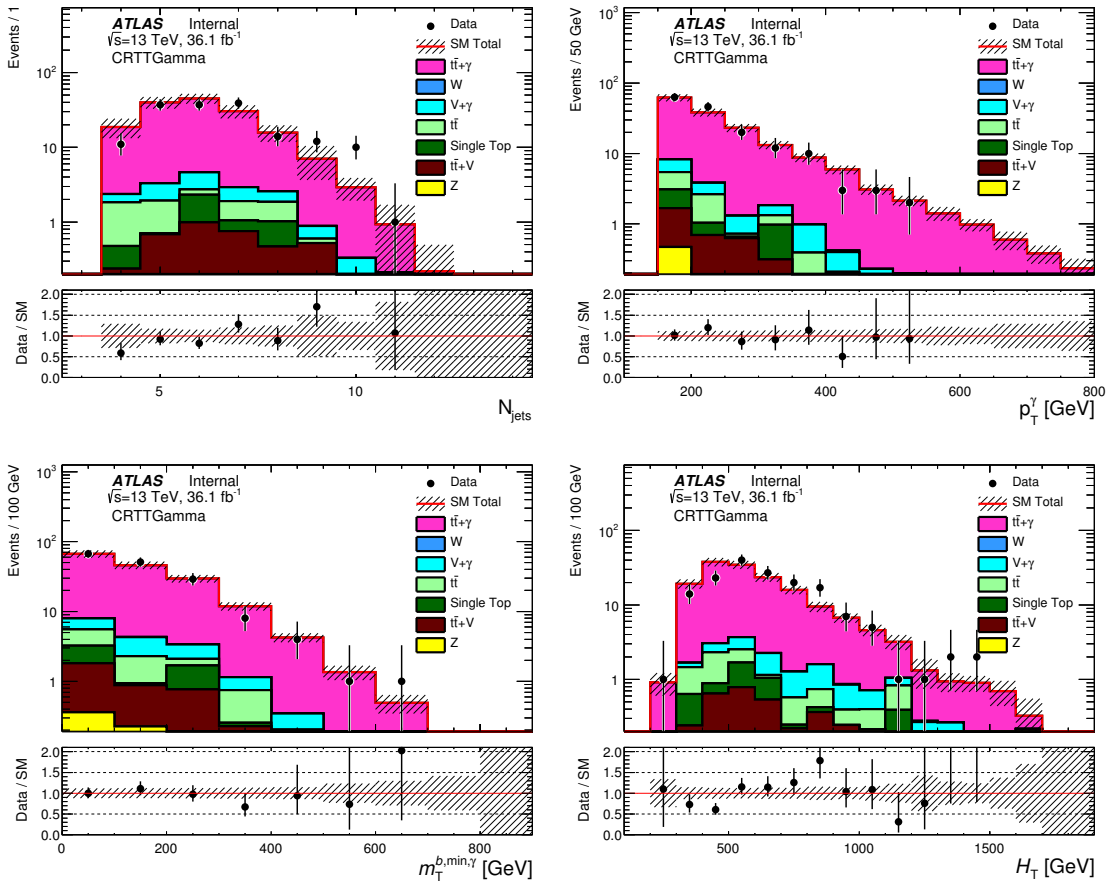

Figure 6.25: Distributions of N_{jets}, γ p_T , $m_T^{b,\min\prime}$ (where the γ p_T is added to the E_T^{miss}) and H_T in CRTTGamma, for 36.07 fb^{-1} of data. Histograms are post-simultaneous fit, normalising simulation to data. The hashed area shows the MC statistical uncertainty only.

would be useful, however, to cross-check this technique or even to directly estimate that contribution. Obviously in practice this is not viable, as the uncertainty on the missing energy would be too large, but similarly to the $t\bar{t}$ + γ control region, the photon from γ + jets decays can be used as a proxy for the Z boson [102]. This alternative method has a couple of advantages over the nominal, charged-lepton method, namely an increase in statistics and a potentially lower systematic uncertainty in the extrapolation from CR to SR. An early study was performed into this method for this analysis, based on the procedure described in the 2016 $\tilde{b}_1$ pair production analysis paper [84].

Table 6.19 describes the selection criteria for this alternative control region, CRZGamma. The trigger requirements are the same as the 2015 trigger requirements for the $t\bar{t}$ + γ control region. The recal-

culated missing energy, $E_T^{\text{miss}\prime}$, is found by treating the γ as invisible, adding its p_T to the E_T^{miss}. The E_T^{miss} itself is required to be low, < 50 GeV, to reject $Z \to \nu\nu$ decays and $W \to l\nu$ decays where the lepton is misreconstructed as a jet. In this low E_T^{miss} region, the missing energy is mostly caused by the jet energy being mismeasured either at hardware level or at reconstruction level, both of which are well-understood by ATLAS [93, 103].

Figure 6.26 shows distributions of several important variables. The data used here represents 3.2 fb^{-1} luminosity from 2015 only, and the MC has been normalised to that luminosity - the simultaneous fit done in other control regions is not used here. The agreement between data and MC is good, in that the data-MC ratio in each variable is flat across all bins. However, a flat factor of ~ 1.3 difference exists between the two, and therein lies the problem with this method. The simulation used to model γ + jets is leading-order only, which works well at low jet multiplicities, for example in the ATLAS 2016 $\tilde{b}_1$ pair production analysis [84], but in a region with high jet multiplicity as is used here, modelling the additional jets becomes problematic. In a control region that enhances that background source, the mis-modelling due to missing higher-order loop processes becomes obvious. This causes the theoretical systematic uncertainty for the background to become large, which counters the reduced uncertainty from the extrapolation, and the point of using this alternate method is rendered moot. As a result, the nominal charged-lepton estimation method was chosen as the default method for estimating Z + jets.

Selection	CRZGamma
Trigger	`HLT_g120_loose`
Jet multiplicity	≥ 4
$N_{\text{b-tag}}$	≥ 1
N_γ	Exactly 1
N_{lep}	Exactly 0
Jet p_T	(80, 80, 40, 40) GeV
Photon p_T	> 130 GeV
E_T^{miss}	< 50 GeV
$E_T^{\text{miss}\prime}$	> 70 GeV

Table 6.19: Selection criteria for the unused CRZGamma. $E_T^{\text{miss}\prime}$ is the recalculated E_T^{miss} where the γ p_T is added to the E_T^{miss} to emulate Z to invisible decays.

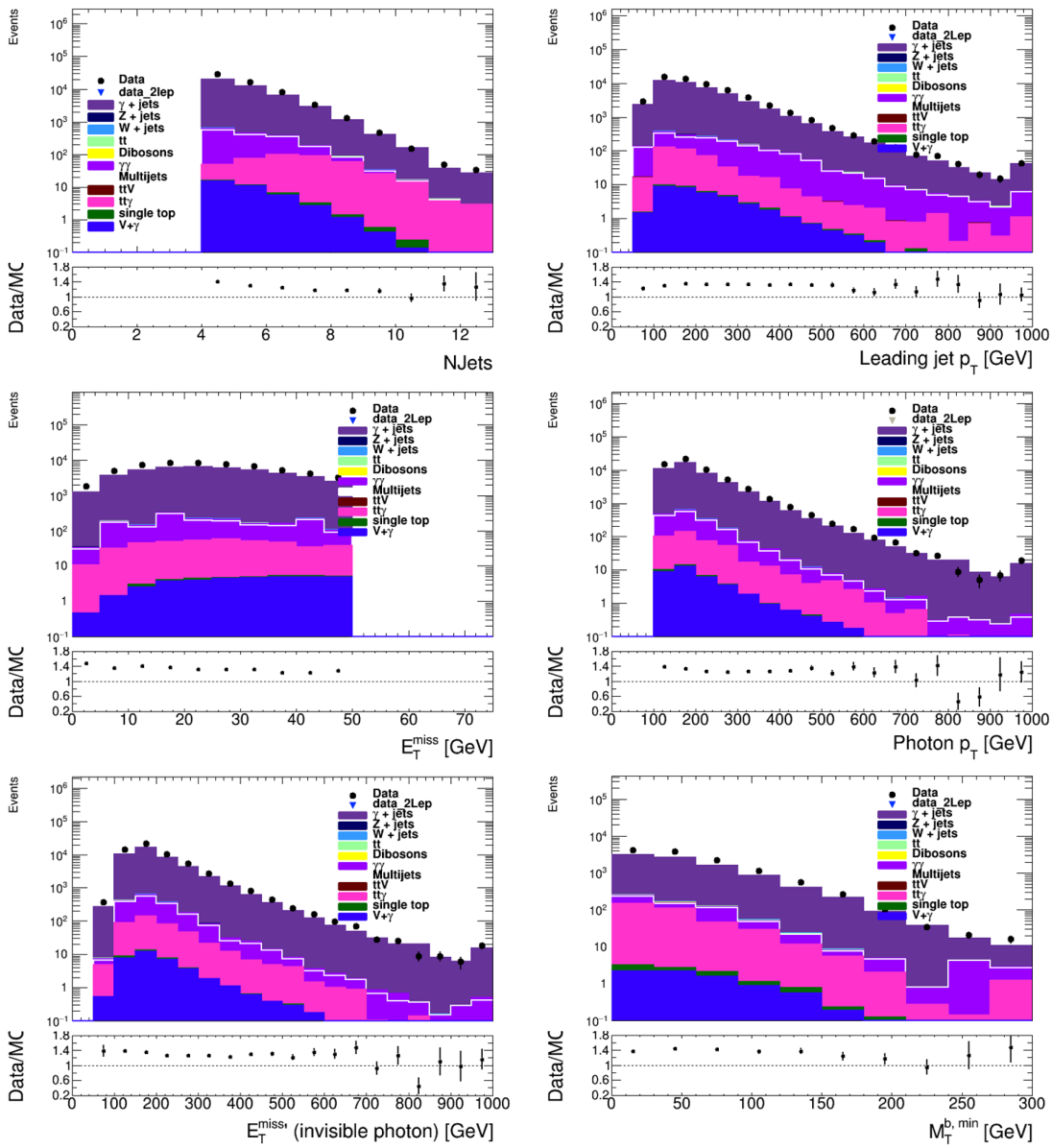

Figure 6.26: Selection of variables for the unused CRZGamma. *From left to right, top to bottom*: N_{jets}, leading jet-p_T, E_T^{miss}, p_T^{γ}, $E_T^{\text{miss}\prime}$ and $m_T^{b,\text{min}\prime}$. The $m_T^{b,\text{min}\prime}$ is calculated using the $E_T^{\text{miss}\prime}$ with the γ p_T added. The MC samples are normalised to 3.2 fb^{-1} luminosity, which is the total amount of data collected in 25 ns bunch spacing pp collisions in 2015. Statistical error on the data is shown, but very small.

6.6 Systematic Uncertainty Evaluation

Several sources of experimental and theoretical systematic uncertainties are considered in the SM background estimates and signal expectations, and are included in the profile likelihood fit described in Section 6.5. These systematic uncertainties can impact the expected event yields in the control and signal regions, as well as the transfer factors used for normalisation during the extrapolation of the background expectation from CR to SR.

The dominant uncertainty to all SRs, except for SRB, is the statistical uncertainty on the estimate of the total background contribution. The main sources of detector-related systematic uncertainties in the SM background estimates originate from the jet energy scale (JES) and resolution (JER). The JES is a calibration or correction applied to jets related to the calorimeter's response to the true jet energy, including uncertainties related to jet flavour composition and pileup, and is derived in bins of p_T and η from different in-situ techniques [104]. Variations up and down are applied to the expected event yields, and the mean shift in yield with respect to the nominal value is taken as the uncertainty. The JER uncertainty is derived as a one-sided variation by comparing data to MC simulation [103], and is calculated by smearing the momentum of every jet in simulated events by this variation, with the mean shift in expected event yield compared to the nominal value taken as the uncertainty. The combined JES+JER uncertainty on the background estimates is 3-19% in SRA and SRB, 7-27% in SRC and 10-11% in SRD.

The uncertainty on the b-tagging efficiency has a large contribution to both signal and background processes, due to the 2 b-jet requirement in the majority of control, validation and signal regions. Scale factor uncertainties in b-tagging are derived based on jet flavour and kinematics, with up and down variations being used to calculate the uncertainties on the b-jet scale factors, as well as corrections for mis-tagging c-jets and light-flavour jets. In total, this uncertainty source contributes uncertainties of 1-3% in SRA, 4-5% in SRB, 1-4% in SRC and 7-9% in SRD.

Lepton reconstructed and identification uncertainties are considered, but have a small impact. All jet- and lepton-related uncertainties are propagated through to the calculation of the E_T^{miss}, given that this depends entirely on the reconstruction of other objects, and additional uncertainties on the energy and resolution of the soft E_T^{miss} term are also included, using data-driven techniques with $Z \to \mu\mu$ events. The uncertainty on the soft term of the E_T^{miss} is most significant in SRC, ranging from 1 to 15%, while contributing less than 4% in other signal regions. An uncertainty due to pileup is also considered, calculated as a two-sided variation, with a contribution in SRA of 5-10%, in SRB of 1-8%, in SRC of 1-14% and SRD of 1-9%.

The uncertainty on the combined 2015 + 2016 integrated luminosity recorded by ATLAS is 3.2%. This is derived from a preliminary calibration of the luminosity scale, using x-y beam-separation scans

performed in August 2015, May 2016 and October 2016.

Theoretical uncertainties in the modelling of the SM background in each SR are evaluated by considering variations with respect to the default settings and choices for the event generation. For each variation considered, the systematic uncertainty is estimated as an uncertainty on the transfer factor used in the extrapolation from SR to CR (Eq. 6.4).

For the W/Z + jets background processes, the modelling uncertainties are evaluated using SHERPA samples, varying up and down the renormalisation and factorisation scales, and the merging and resummation scales. The uncertainty in the transfer factor is given by Eq. 6.5, where X is the systematic variation and *up* and *down* refer to the variation compared to the nominal samples. The resulting impact on the total background yields from the Z + jets (W + jets) theoretical uncertainties are 5-13% (1%) for SRA, 5-7% (1-2%) for SRB, $<$ 1% (1-6%) for SRC and 4-5% (1-3%) for SRD.

$$\Delta_X = \frac{T_f^{\text{up}} - T_f^{\text{down}}}{T_f^{\text{up}} + T_f^{\text{down}}} \tag{6.5}$$

For the $t\bar{t}$ background, uncertainties are evaluated due to the hard scattering generation (comparing `MadGraph5AMC@NLO` with `Powheg-Box`, using Eq. 6.6), the choice of the parton showering model (`PYTHIA` against `Herwig++`, using Eq. 6.7) and the emission of additional partons in the initial and final states (varying up and down the amount of radiation, evaluated with Eq. 6.5). The largest impact of $t\bar{t}$ systematics is 17-75% in SRC, going down to 5-20% in the other signal regions. In general, evaluation of these systematics is limited by MC statistics in the variation samples, which results in a larger uncertainty than would otherwise be found.

$$\Delta_{\text{hard scatter}} = \frac{T_f^{\text{PowHeg}} - T_f^{\text{MC@NLO}}}{T_f^{\text{PowHeg}}} \tag{6.6}$$

$$\Delta_{\text{PS}} = \frac{T_f^{\text{PYTHIA}} - T_f^{\text{HERWIG++}}}{T_f^{\text{PYTHIA}}} \tag{6.7}$$

The effect of $t\bar{t} + Z$ theory uncertainties on the yields of that process is determined by performing the analysis with particle-level variations. The systematic uncertainty is calculated as the difference in yields due to the variations, relative to the nominal values. Scale variations (up and down by a factor of two) and variations in parton distribution function are used to estimate the uncertainty due to different renormalisation and factorisation scales. The difference in cross-section between the `Sherpa+OpenLoops` and `MadGraph_aMC@NLO` cross-section ratios is summed in quadrature with the scale

and PDF variations. This leads to uncertainties of 3-8% in SRA, 2-5% in SRB, $< 1\%$ in SRC and 2% in SRD.

Theory uncertainties related to single-top backgrounds, dominated by the Wt subprocess, arise from the choice of parton showering model and the level of initial and final state radiation. Both are evaluated as uncertainties on the transfer factor. The uncertainty from parton showers is evaluated by comparing the nominal `PowHegPYTHIA` tuning to a `PowHegHERWIG++` tuning variation, using Eq. 6.7. Uncertainties from ISR/FSR are evaluated by varying up and down the level of radiation for both top and antitop final states, using Eq. 6.5. This leads to uncertainty values of 6-11% in SRA and SRB, 1-6% in SRC and 10-12% in SRD. An additional 30% uncertainty is applied to account for the effect of interference between single-top and $t\bar{t}$ production.

Signal-related systematic uncertainties are only taken into account during the limit-setting process. These arise mostly from detector and acceptance effects, and are calculated in the same way as for the background contributions. Each signal model and mass point has a different set of uncertainties, hence only the range of values is stated here. The JES and JER are in general the largest contributors, ranging from 3 to 23% and 6 to 16% respectively. The uncertainty due to pileup corrections ranges from 6 to 20%. Any uncertainty on the estimation of the number of signal events due to cross-section uncertainties is taken into account by calculating two additional limits, corresponding to a $\pm 1\sigma$ change in the predicted cross-section. The cross-section uncertainty itself is $\sim 15\%$.

Tables 6.20 and 6.21 summarise the dominant ($> 1\%$ for at least one SR) systematics for each signal region.

6.7 Results

The observed event yields in data are compared to the total number of expected background events in Tables 6.22 through 6.25. The total background estimate is determined from the simultaneous fit based on the procedure described in Section 6.5. Distributions of discriminating variables are shown in Figures 6.27 through 6.30, for each signal region, with the data overlaid with the simulated SM background and benchmark signal distributions. In these, the background expectations are normalised to the values determined from the simultaneous fit, and the signal is normalised to the total recorded luminosity, 36.07 fb^{-1}. The statistical error on the data is assumed to be Poissonian - in the cases where the number of observed events is very low (roughly 10 or less per bin), this results in an asymmetric error, as the Poisson distribution itself becomes asymmetric, with a short lower tail squeezed between the peak and zero, and a longer upper tail. Hence, some error bars in Figures 6.27-6.31 may appear larger than one would naively expect from a counting experiment.

	SRA-TT	SRA-TW	SRA-T0	SRB-TT	SRB-TW	SRB-T0
Pileup	10	5	5	8	1	3
JER	10	12	4	3	4	3
$t\bar{t}$ theory	10	6	3	10	11	12
MC statistical	10	8	8	9	4	3
Single top theory	6	3	5	3	4	5
μ_Z	6	10	7	5	6	4
$\mu_{t\bar{t}+Z}$	6	3	2	4	3	2
$\mu_{\text{single top}}$	5	3	5	4	4	5
JES	4	7	1	7	4	<1
$t\bar{t}$ theory	2	<1	<1	1	<1	<1
E_T^{miss} soft term	2	2	<1	1	<1	<1
b-tagging	1	3	2	5	4	4
Z theory	1	3	2	<1	1	<1
Multijet estimate	1	<1	<1	2	2	<1
μ_W	1	1	1	2	1	2
Diboson theory	<1	2	<1	<1	<1	<1
$\mu_{t\bar{t}}$	<1	<1	<1	2	2	1

Table 6.20: Dominant systematic uncertainties ($> 1\%$ in at least one SR) for SRA and SRB. Values are given as percentages relative to the total background estimates. The uncertainty due to the normalisation from a control region for a given SR and background are indicated by μ_p for each process p. The MC statistical error is also given, for reference.

No significant excess above the SM expectation is observed in any of the signal regions; the p-values, which express the probability that the background fluctuates to the data or above, and the model independent limits are shown in Table 6.26. The smallest p-values, under the background-only hypothesis, are 0.27 for both SRB-T0 and SRD-high, and 0.29 in SRA-TT. In cases where the data fluctuate below the background expectation, the p-value is set to 0.50, to prevent a statistical under-fluctuation from misleadingly lowering the p-value. The 95% confidence level (CL) upper limits on the number of BSM events in each signal region are derived using the CL_s prescription [73] and calculated from asymptotic formulae [72]. Model independent limits are placed on the highest allowed BSM cross-sections, given the number of events observed, defined as $\sigma_{\text{vis}} = \sigma \cdot A \cdot \epsilon$, where σ is the production cross-section, A is the acceptance and ϵ is the selection efficiency for a BSM signal.

The detector acceptance multiplied by the efficiency ($A \cdot \epsilon$) is calculated for several signal regions and their benchmark points, by finding the percentage of signal events in the benchmark sample that are

	SRC1	SRC2	SRC3	SRC4	SRC5	SRD-low	SRD-high
$t\bar{t}$ theory	27	11	14	11	71	12	10
Multijet estimate	12	3	<1	<1	<1	2	2
MC statistical	6	5	5	7	23	13	6
$\mu_{t\bar{t}}$	4	6	6	5	5	1	1
JES	4	5	2	2	17	8	4
JER	4	10	6	5	10	3	6
Single top theory	3	2	2	3	<1	5	6
$\mu_{\text{single top}}$	3	2	2	3	<1	5	6
b-tagging	2	2	<1	2	4	9	7
E_T^{miss} soft term	1	3	2	3	15	4	3
Pileup	<1	1	<1	2	14	9	<1
μ_W	<1	<1	1	3	3	3	1
W theory	<1	<1	1	3	2	<1	<1
μ_Z	<1	<1	<1	<1	<1	4	5
$\mu_{t\bar{t}+Z}$	<1	<1	<1	<1	<1	2	2
Z theory	<1	<1	<1	<1	<1	<1	<1
$t\bar{t}$ theory	<1	<1	<1	<1	<1	<1	<1

Table 6.21: Dominant systematic uncertainties (> 1% in at least one SR) for SRC and SRD. Values are given as percentages relative to the total background estimates. The uncertainty due to the normalisation from a control region for a given SR and background are indicated by μ_p for each process p. The MC statistical error is also given, for reference.

rejected as a result of the object definitions and the SR definition. The $A \cdot \epsilon$ for SRA is 9% for the signal benchmark point $m_{\tilde{t}_1} = 1000$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV, while SRB has an $A \cdot \epsilon$ of 1.4% for the $m_{\tilde{t}_1} = 600$ GeV, $m_{\tilde{\chi}_1^0} = 300$ GeV benchmark model. For SRD-low and SRD-high, the $A \cdot \epsilon$ is 0.05% and 0.5% for the $m_{\tilde{t}_1} = 400$ GeV, $m_{\tilde{\chi}_1^\pm} = 100$ GeV, $m_{\tilde{\chi}_1^0} = 50$ GeV and $m_{\tilde{t}_1} = 700$ GeV, $m_{\tilde{\chi}_1^\pm} = 200$ GeV, $m_{\tilde{\chi}_1^0} = 100$ GeV benchmark models, assuming a 100% branching fraction for $\tilde{t}_1 \to b\tilde{\chi}_1^\pm$ decays. Finally, combining the R_{ISR} windows in SRC gives an $A \cdot \epsilon$ of 0.08% for the $m_{\tilde{t}_1} = 400$ GeV, $m_{\tilde{\chi}_1^0} = 227$ GeV signal point.

The profile likelihood ratio test statistic is used to set limits on the direct pair production of stops. The signal strength parameter is allowed to float in the fit [100], and any signal contamination in the CRs is taken into account. Again, limits are derived using the CL_s prescription and calculated from asymptotic formulae. The top reconstruction categories in SRA and SRB, and the bins of R_{ISR} in SRC, are statistically combined by multiplying their likelihood functions, which is possible due

	SRA-TT	SRA-TW	SRA-T0
Observed	11	9	18
Fitted background events			
Total SM	8.6 ± 2.5	9.4 ± 2.9	18.7 ± 5.4
$t\bar{t}$	$0.70^{+0.89}_{-0.70}$	$0.51^{+0.55}_{-0.51}$	1.31 ± 0.71
W + jets	0.82 ± 0.17	0.89 ± 0.56	2.00 ± 0.96
Z + jets	2.5 ± 1.6	4.9 ± 2.6	9.8 ± 3.9
$t\bar{t} + Z$	3.17 ± 0.78	1.85 ± 0.49	2.61 ± 0.64
Single top	$1.2^{+1.4}_{-1.2}$	$0.70^{+0.76}_{-0.70}$	$2.9^{+3.1}_{-2.9}$
Dibosons	$--$	0.35 ± 0.26	$--$
Multijets	0.21 ± 0.10	0.14 ± 0.09	0.12 ± 0.07

Table 6.22: Expected and observed yields for SRA for 36.07 fb^{-1} total integrated luminosity. The uncertainties include MC statistical, detector systematics and theoretical uncertainties on the extrapolation from CR to SR.

to their orthogonality. In SRD, where the low and high categories overlap, the subregion with the smallest expected 95% CL$_s$ value is chosen for each signal model. Once the signal subregions are combined or chosen, the signal region with the smallest expected 95% CL$_s$ value is chosen from SRA, SRB and SRC for each signal model in the $\tilde{t}_1$-$\tilde{\chi}_1^0$ signal grid. The nominal event yield in each SR is set to the background expectation to determine the expected limits; contours that correspond to $\pm 1\sigma$ uncertainties in the background estimates (σ_{exp}) are also evaluated. The observed event yields determine the observed limits for each SR; these are evaluated for the nominal signal cross sections as well as for $\pm 1\sigma$ theory uncertainties on those cross sections $\sigma^{\text{SUSY}}_{\text{theory}}$.

Figure 6.32 shows the observed (solid red line) and expected (dashed blue line) exclusion limits at 95% CL in the $\tilde{t}_1$-$\tilde{\chi}_1^0$ mass plane for 36.07 fb^{-1} integrated luminosity, for SRA, SRB and SRC. The data excludes $\tilde{t}_1$ masses in the range 400-940 GeV for $\tilde{\chi}_1^0$ masses below 160 GeV, extending the Run 1 limits by 220 GeV. Additional constraints are set in the case when $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0) \approx m_t$, for which $\tilde{t}_1$ masses in the range 250-540 GeV are excluded. With careful inspection, one can note that there is a very small region around $m_{\tilde{t}_1} = m_{\text{top}}$ that is not excluded - this is purely down to the fact that it is impossible to distinguish SUSY events from SM $t\bar{t}$ events in that particular scenario, as the two are essentially identical.

Mixed decay modes, where the $\tilde{t}_1$ is allowed to decay via either the $\tilde{t}_1 \to t\tilde{\chi}_1^0$ process or the $\tilde{t}_1 \to b\tilde{\chi}_1^\pm \to bW^{(*)}\tilde{\chi}_1^0$ process, are considered by combining the results from all four signal regions SRA,

	SRB-TT	SRB-TW	SRB-T0
Observed	38	53	206
Fitted background events			
Total SM	39.4 ± 9.2	52 ± 11	179 ± 40
$t\bar{t}$	7.3 ± 4.3	12.4 ± 5.7	43 ± 22
W + jets	7.8 ± 2.9	4.8 ± 1.3	25.8 ± 8.9
Z + jets	9.0 ± 4.2	16.8 ± 6.8	60 ± 20
$t\bar{t} + Z$	9.3 ± 2.1	10.8 ± 2.0	20.6 ± 4.2
Single top	$4.2 \, {}^{+\,4.5}_{-\,4.2}$	$5.9 \, {}^{+\,6.3}_{-\,5.9}$	$26 \, {}^{+\,28}_{-\,26}$
Dibosons	0.13 ± 0.07	0.60 ± 0.43	1.04 ± 0.73
Multijets	1.54 ± 0.64	1.01 ± 0.88	1.8 ± 1.5

Table 6.23: Expected and observed yields for SRB for 36.07 fb^{-1} total integrated luminosity. The uncertainties include MC statistical, detector systematics and theoretical uncertainties on the extrapolation from CR to SR.

SRB, SRC and SRD. The exclusion limits for this are shown in Figure 6.33, under the assumptions of $\mathcal{B}(\tilde{t}_1 \to t\tilde{\chi}_1^0) = 0\%, 25\%, 50\%,$ or 75%. This scenario excludes $\tilde{t}_1$ masses between 450 and 850 GeV, for $m_{\tilde{\chi}_1^0} < 240$ GeV and $\mathcal{B}(\tilde{t}_1 \to t\tilde{\chi}_1^0) = 50\%$.

	SRC1	SRC2	SRC3	SRC4	SRC5
Observed	20	22	22	1	0
Fitted background events					
Total SM	20.6 ± 6.6	27.5 ± 4.9	18.9 ± 3.5	7.7 ± 1.4	0.90 ± 0.71
$t\bar{t}$	12.8 ± 5.9	22.1 ± 4.2	14.6 ± 3.2	4.92 ± 0.98	$0.63 \, ^{+\,0.69}_{-\,0.63}$
W + jets	0.81 ± 0.37	1.93 ± 0.48	1.91 ± 0.63	1.93 ± 0.45	0.21 ± 0.12
Z + jets	$--$	$--$	$--$	$--$	$--$
$t\bar{t} + Z$	0.29 ± 0.18	0.59 ± 0.38	0.56 ± 0.31	0.08 ± 0.08	0.06 ± 0.03
Single top	$1.7 \, ^{+\,2.0}_{-\,1.7}$	$1.2 \, ^{+\,1.8}_{-\,1.2}$	$1.2 \, ^{+\,1.4}_{-\,1.2}$	$0.72 \, ^{+\,0.77}_{-\,0.72}$	$--$
Dibosons	0.39 ± 0.33	0.21 ± 0.11	0.29 ± 0.18	$--$	$--$
Multijets	4.6 ± 2.4	1.58 ± 0.77	0.32 ± 0.17	0.04 ± 0.02	$--$

Table 6.24: Expected and observed yields for SRC1-5 for 36.07 fb^{-1} total integrated luminosity. The uncertainties include MC statistical, detector systematics and theoretical uncertainties on the extrapolation from CR to SR.

	SRD-low	SRD-high
Observed	27	11
Fitted background events		
Total SM	25.2 ± 7.7	8.6 ± 2.4
$t\bar{t}$	3.2 ± 3.2	0.98 ± 0.89
W + jets	6.1 ± 3.0	1.06 ± 0.34
Z + jets	6.9 ± 3.0	3.2 ± 1.3
$t\bar{t} + Z$	4.0 ± 1.0	1.38 ± 0.39
Single top	$3.9 \, ^{+\,4.2}_{-\,3.9}$	$1.5 \, ^{+\,1.6}_{-\,1.5}$
Dibosons	$--$	$--$
Multijets	1.12 ± 0.37	0.40 ± 0.15

Table 6.25: Expected and observed yields for SRD for 36.07 fb^{-1} total integrated luminosity. The uncertainties include MC statistical, detector systematics and theoretical uncertainties on the extrapolation from CR to SR.

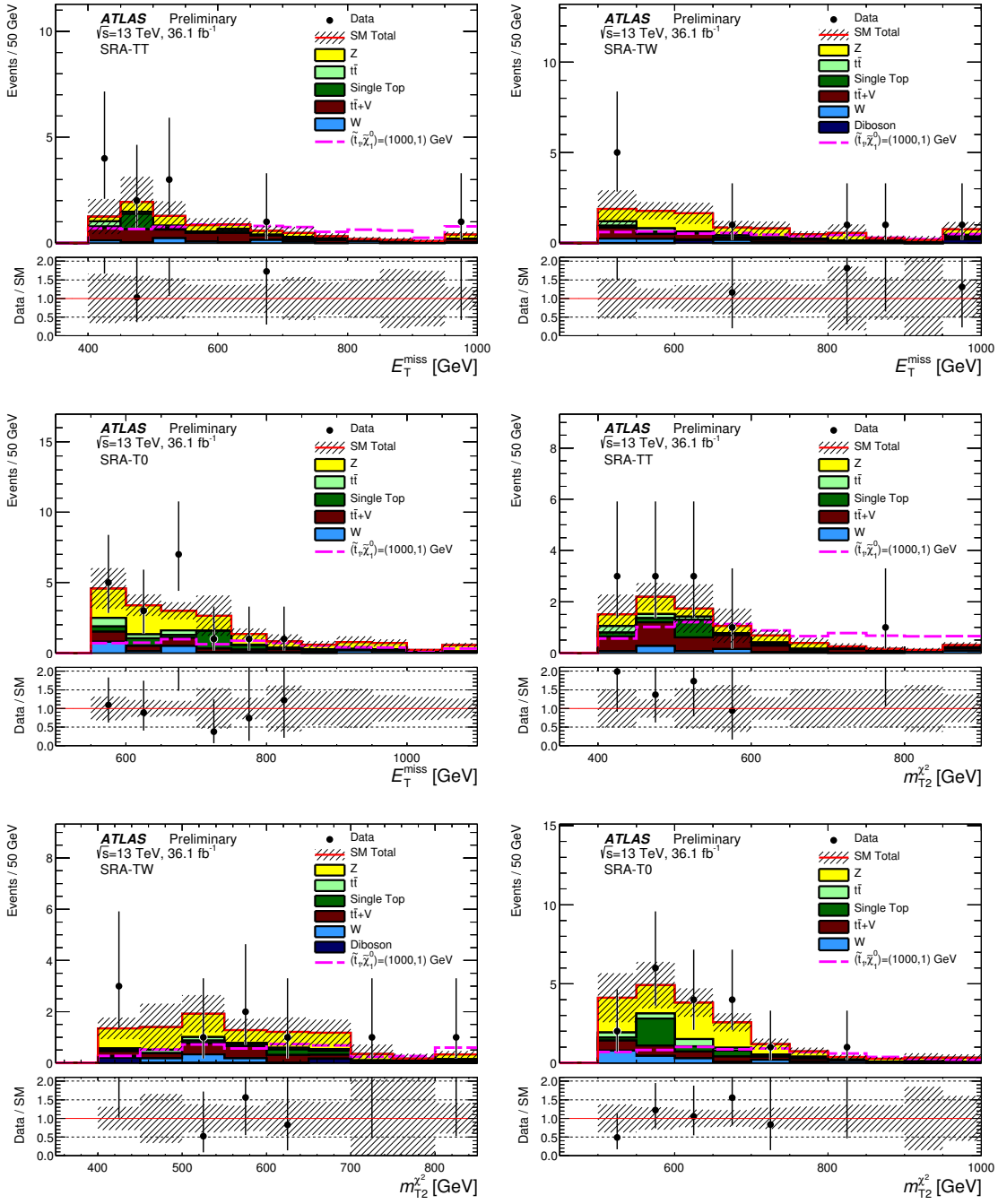

Figure 6.27: Distributions of E_T^{miss} and m_{T2} for the three SRA top reconstruction categories using 36.07 fb^{-1} of data. The stacked histograms show the SM expectation, and the hashed uncertainty band around the SM expectation shows the MC statistical and detector-related systematic uncertainties.

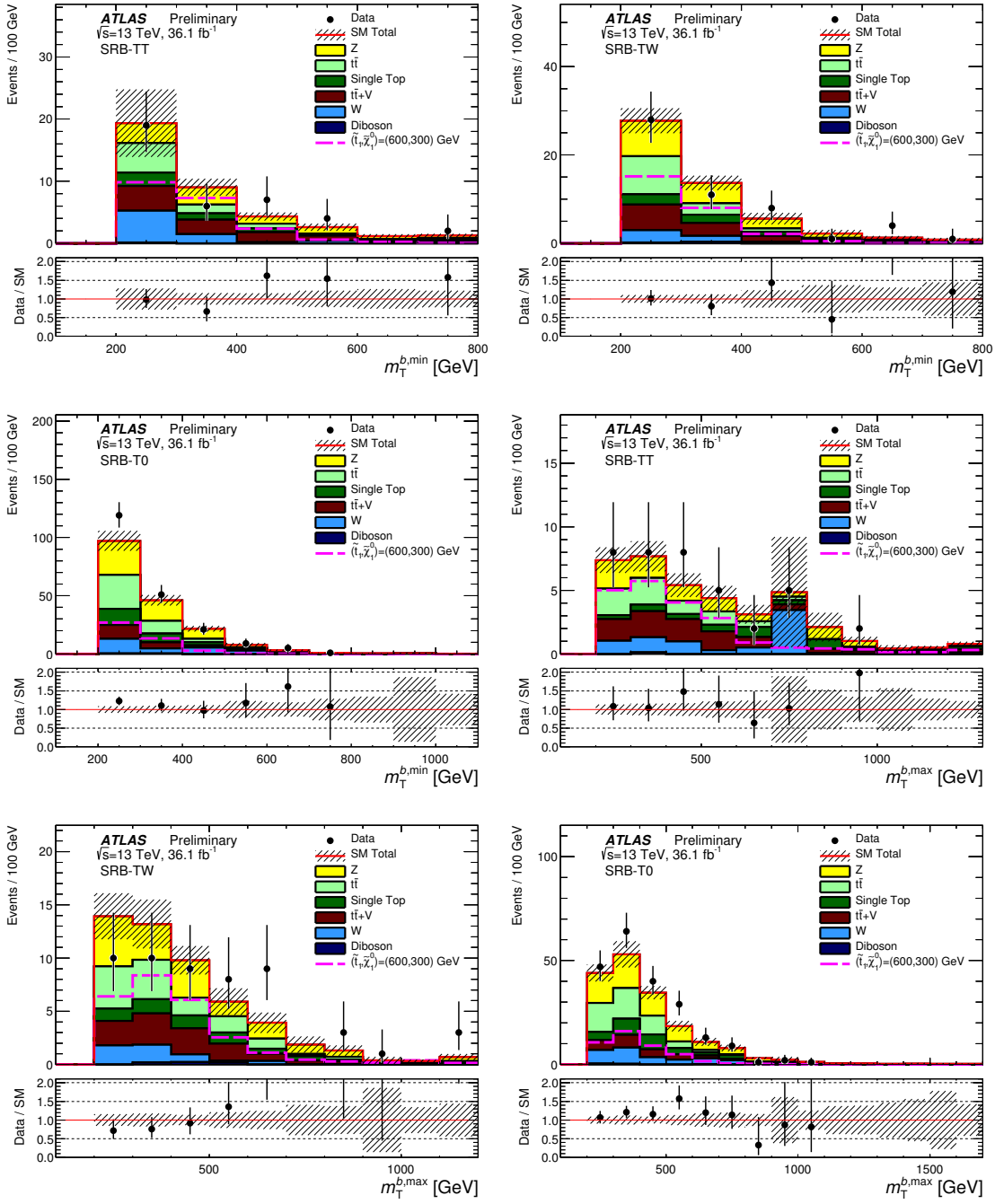

Figure 6.28: Distributions of $m_T^{b,\min}$ and $m_T^{b,\max}$ for SRB in all three top reconstruction categories, using 36.07 fb^{-1} of data. The stacked histograms show the SM expectation, and the hashed uncertainty band around the SM expectation shows the MC statistical and detector-related systematic uncertainties.

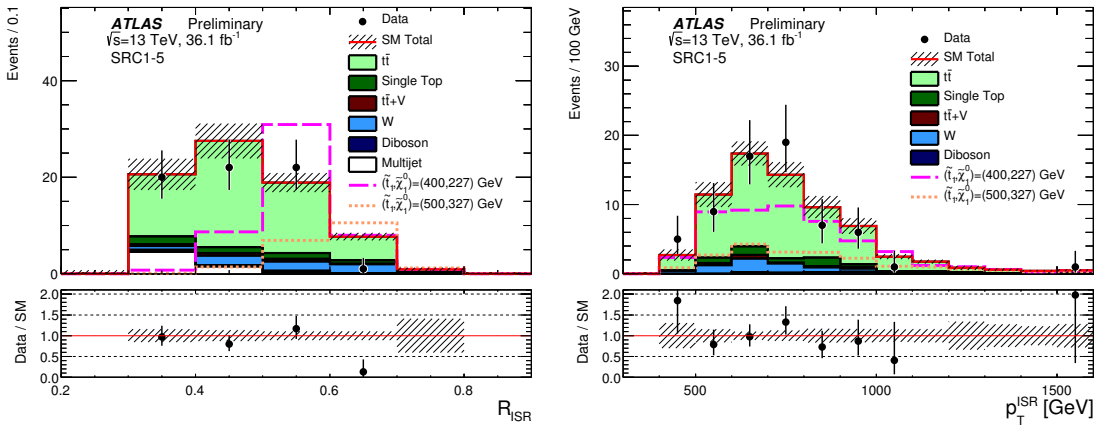

Figure 6.29: Distrbution of R_{ISR} and p_T^{ISR} for SRC1-5 combined, using 36.07 fb^{-1} of data. The stacked histograms show the SM expectation, and the hashed uncertainty band around the SM expectation shows the MC statistical and detector-related systematic uncertainties.

Signal channel	$\langle\epsilon\sigma\rangle^{95}_{\mathrm{obs}}$[fb]	S^{95}_{obs}	S^{95}_{exp}	$p(s=0)$ (σ)
SRA-TT	0.31	11.2	$9.1^{+4.3}_{-2.5}$	0.29 (0.57)
SRA-TW	0.25	9.0	$9.2^{+3.6}_{-2.8}$	0.50 (0.00)
SRA-T0	0.40	14.6	$14.9^{+5.2}_{-4.3}$	0.50 (0.00)
SRB-TT	0.65	23.4	$24.0^{+7.8}_{-6.8}$	0.50 (0.00)
SRB-TW	0.73	26.2	$26.0^{+8.8}_{-6.6}$	0.48 (0.05)
SRB-T0	2.93	106	91^{+24}_{-22}	0.27 (0.61)
SRC1	0.44	16.0	$16.3^{+5.8}_{-4.2}$	0.50 (0.00)
SRC2	0.35	12.6	$15.5^{+5.9}_{-4.2}$	0.50 (0.00)
SRC3	0.44	15.8	$12.8^{+4.7}_{-2.7}$	0.30 (0.54)
SRC4	0.09	3.1	$6.5^{+3.3}_{-2.1}$	0.50 (0.00)
SRC5	0.06	3.0	$3.9^{+1.0}_{-0.3}$	0.50 (0.00)
SRD-low	0.57	20.7	$19.8^{+6.8}_{-4.9}$	0.43 (0.19)
SRD-high	0.32	11.6	$9.6^{+3.9}_{-2.4}$	0.27 (0.60)

Table 6.26: *Left to right*: 95% CL upper limits on the visible cross section ($\langle\epsilon\sigma\rangle^{95}_{\mathrm{obs}}$) and on the number of signal events (S^{95}_{obs}). The third column (S^{95}_{exp}) shows the 95% CL upper limit on the number of signal events, given the expected number (and $\pm 1\sigma$ excursions on the expectation) of the background events. The last column indicates the discovery p-value ($p(s=0)$) and the significance for the p-value (σ). The p-value is set to 0.50 when the observed event yield is less than the expected event yield.

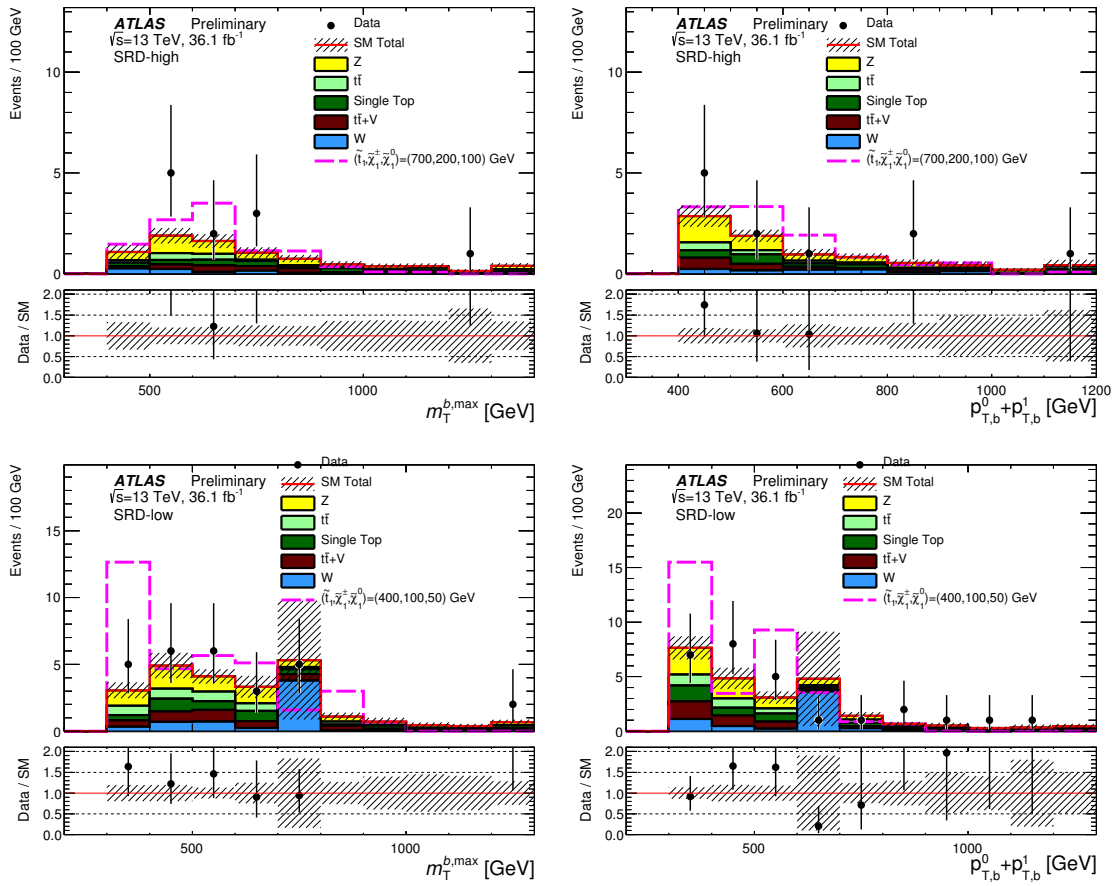

Figure 6.30: Distribution of $m_T^{b,\max}$ and b-jet $p_T^1 + p_T^2$ for SRD-high and SRD-low, using 36.07 fb^{-1} of data. The stacked histograms show the SM expectation, and the hashed uncertainty band around the SM expectation shows the MC statistical and detector-related systematic uncertainties.

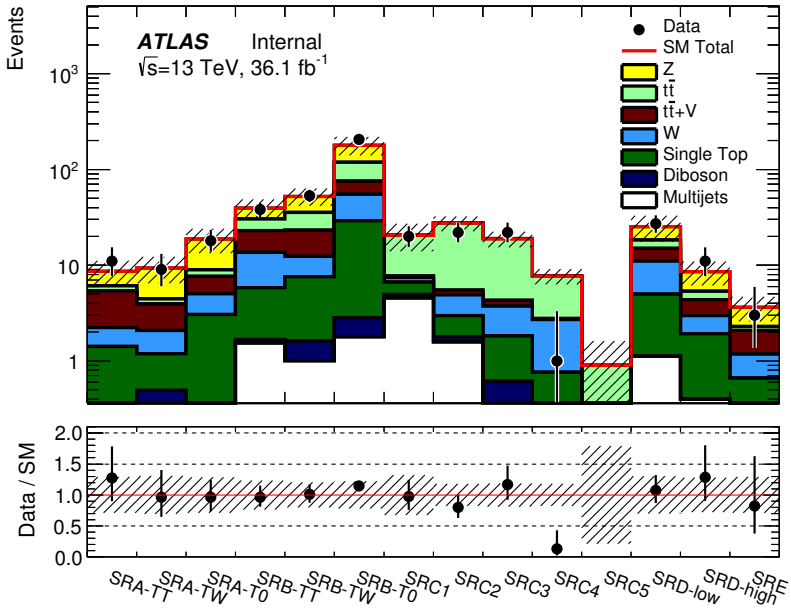

Figure 6.31: Final yields for all signal regions, with 36.07 fb^{-1} of data. The stacked histograms show the SM expectation and the hatched uncertainty band around the SM expectation shows the MC statistical and detector-related systematic uncertainties. An additional signal region, SRE, is included here, but is detailed in the following chapter as it relates to a reinterpretation of the data used in this analysis.

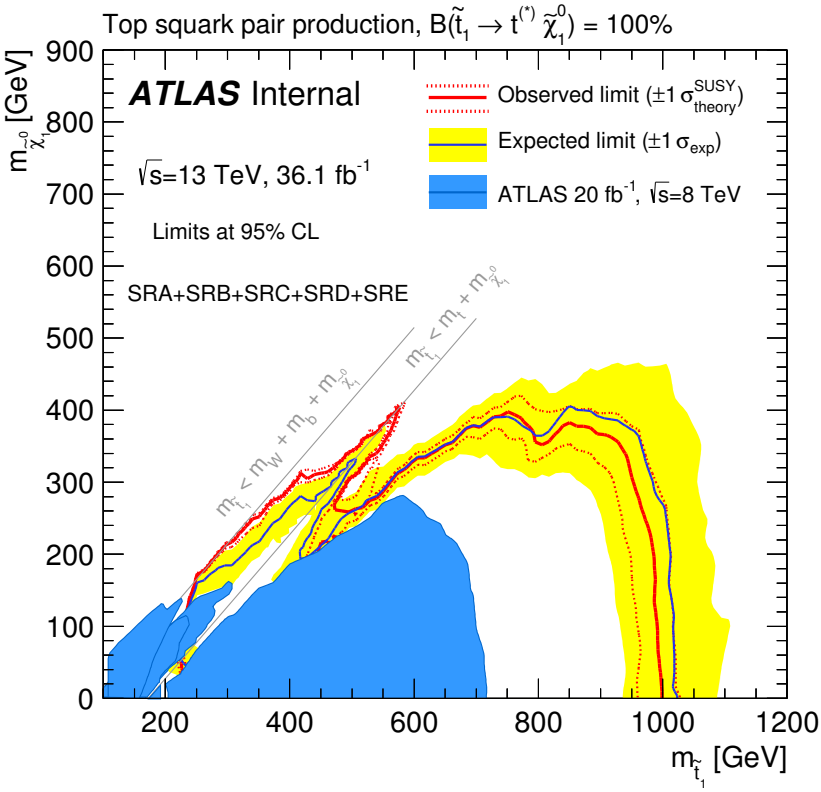

Figure 6.32: Expected (blue line) and observed (red line) exclusion limits at 95% CL as a function of $\tilde{t}_1$ and $\tilde{\chi}_1^0$ masses in the scenario where both top squarks decay via $\tilde{t}_1 \to t\tilde{\chi}_1^0$ and the top quark decays hadronically. Uncertainty bands corresponding to the $\pm 1\sigma$ variation on the expected limit (yellow band) and the sensitivity of the observed limit to $\pm 1\sigma$ variations of the signal theoretical uncertainties (red dotted lines) are also indicated.

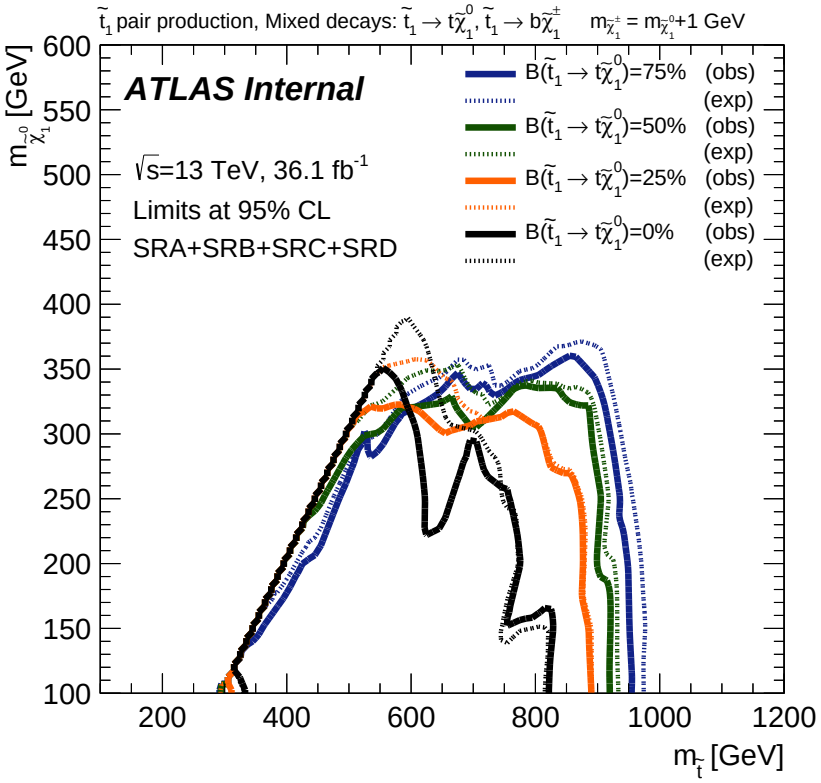

Figure 6.33: Expected (dashed lines) and observed (solid lines) exclusion limits at 95% CL as a function of $\tilde{t}_1$ and $\tilde{\chi}_1^0$ masses in the scenario where both top squarks are allowed to decay via $\tilde{t}_1 \to t\tilde{\chi}_1^0$ or $\tilde{t}_1 \to b\tilde{\chi}_1^\pm \to bW^{(*)}\tilde{\chi}_1^0$. Different curves are shown for the scenarios in which the $\tilde{t}_1$ decays via $\tilde{t}_1 \to t\tilde{\chi}_1^0$ with 0%, 25%, 50% or 75% probability.

6.8 Summary

The results from the search for top squark production based on 36.07 fb^{-1} data of $\sqrt{s} = 13$ TeV pp collisions recorded by the ATLAS experiment in 2015 and 2016. Top squarks are searched for in final states with high jet multiplicity, high-p_T jets and large missing transverse momentum. The top squark is assumed to decay via $\tilde{t}_1 \to t\tilde{\chi}_1^0$ with large or small values of $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$, and via $\tilde{t}_1 \to b\tilde{\chi}_1^\pm \to bW^{(*)}\tilde{\chi}_1^0$. The SM background is predicted by MC simulation, and data in several specially-designed signal regions are compared to the SM prediction.

No significant excess above the expected SM background prediction is observed. Exclusions limits at 95% confidence level on the combination of top squark and $\tilde{\chi}_1^0$ mass are derived, with $m_{\tilde{t}_1}$ excluded up to 940 GeV for $\tilde{\chi}_1^0$ masses below 160 GeV in the nominal signal channel $\tilde{t}_1 \to t\tilde{\chi}_1^0$. Additionally, independent limits and p-values for each signal region are reported, with the smallest p-values being 27% in both SRB-T0 and SRD-high. The limits placed here significantly extend previous results. While 'natural' SUSY could still be found, with sub-TeV $\tilde{t}_1$ mass, the phase-space where this is possible is being increasingly limited - either more technical search methods must be implemented, or more complex decay models should be investigated, in order to maintain compatibility with the idea of naturalness.

Chapter 7

Reinterpretations and Future Work

As the LHC has reached its maximum collision energy, there will not be an increase of $\sqrt{s}$ for a very long time. For BSM theories that do predict TeV-scale massive particles, the lack of evidence so far for any new physics puts them into an awkward position (as demonstrated by the limits shown in Figure 7.1) [105]. However, some theories - like Supersymmetry, as one example - still have room to be found, and the continuing in-take of new data from the LHC can provide evidence for their existence. The focus on simple, easily-modelled signals is starting to fade as a result, as physicists start to work out how to eke the most out of the data they have; and even with those simplified models traditionally used for SUSY searches, experimental techniques are starting to evolve away from basic cut-and-count analyses, such as the one in Chapter 6. More complex search techniques may be more difficult to put into practice, but may provide higher efficiencies than would otherwise be possible.

This chapter details three of these potential evolutions. Sections 7.1 and 7.2 describe reinterpretations of the hadronic stop data used in Chapter 6, a gluino-mediated stop pair production model and a direct dark matter production model respectively. Section 7.3 details a study testing the potential of multivariate analysis techniques to improve the sensitivity to compressed $\tilde{t}_1$ decay scenarios, where the mass difference between the $\tilde{t}_1$ and the $\tilde{\chi}_1^0$ is very close or equal to the top quark mass. This final state is very tricky to describe using cut-and-count methods alone, as detailed in the previous chapter, but a multivariate technique that can take advantage of smaller differences in any one variable may be more effective.

7.1 Gluino-Mediated Stop Decays

The gluino ($\tilde{g}$, the supersymmetric partner of the gluon) is predicted to have the largest pair production cross-section of all sparticles at the LHC (as demonstrated in Figure 6.1). It has been well-studied

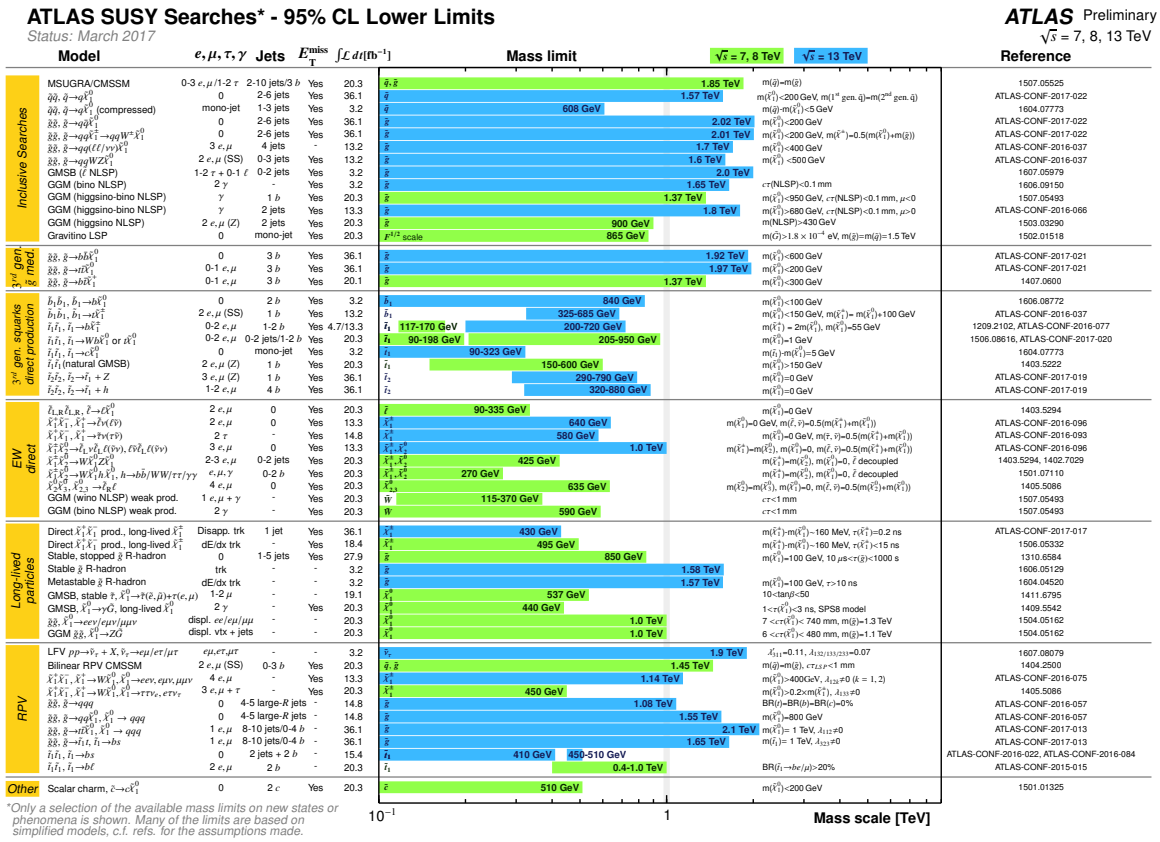

Figure 7.1: Official ATLAS limits on a variety of SUSY particles, from a number of different analysis methods. A number of these limits are already pushing the boundary of what is possible with the LHC, using simple analysis methods. However there is still room for improvement, and new analysis techniques such as multivariate analyses may provide that extra boost in sensitivity.

by ATLAS throughout both Run 1 and Run 2 and strong limits have been placed on its mass [106]. Most gluino searches use simplified models where the gluino either decays directly to the lightest supersymmetric particle, $\tilde{\chi}_1^0$, with a quartic coupling vertex producing two quarks and the $\tilde{\chi}_1^0$, or where it decays to the $\tilde{\chi}_1^0$ via the chargino $\tilde{\chi}_1^\pm$; examples of this can be seen in Figure 7.2.

However, a gluino decay chain can be considered, whereby the gluino decays to the $\tilde{\chi}_1^0$ via a stop, $\tilde{t}_1$, in a process like $\tilde{g} \rightarrow t\tilde{t}_1(\rightarrow q\tilde{\chi}_1^0)$ [107], shown in Figure 7.3. The flavour of the quark produced in the stop decay depends on the mass difference between the $\tilde{t}_1$ and the $\tilde{\chi}_1^0$ - if the difference is large, the most likely quark to be produced is the top. If the mass splitting is small, much less than the top mass, then the quark produced is likely to be a c- or light-flavour jet. More importantly though, if the splitting is very small, $\mathcal{O}(5\text{GeV})$, then the charm jet is too soft to be seen experimentally, lost in the

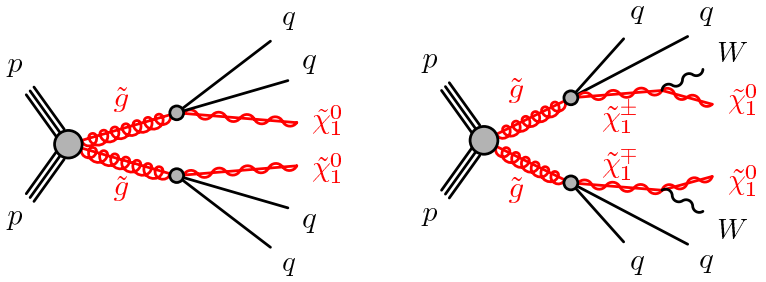

Figure 7.2: Simplified models of gluino decays, where a quartic coupling vertex causes the gluino to decay to two quarks and an electroweak sparticle. This results in a large number of jets in the final state, plus large missing transverse momentum from the $\tilde{\chi}_1^0$.

background QCD interactions. This experimental signature is nearly identical to the final state for direct stop pair production - two top quarks plus missing energy, E_T^{miss}. The same data used in that analysis, described in the previous chapter, can be used to search for these gluino decays. These gluino decays tend to be higher-energy than the nominal $\tilde{t}_1$ decays, however, and several different variables are used here to encapsulate that boosted topology, including the $m^0_{\text{jet},R=0.8}$ and $m^1_{\text{jet},R=0.8}$ variables defined in Section 6.3 but not used in the $\tilde{t}_1$-specific signal regions, as well as the sum of the transverse momentum of all jets, H_T.

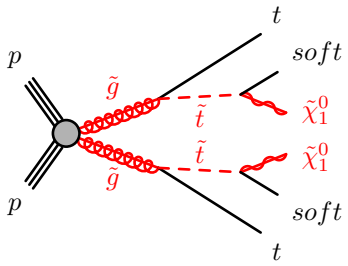

Figure 7.3: Simplified model for gluino-mediated stop decays, used in this reinterpretation of direct stop pair production. The soft jet from the $\tilde{t}_1$ decay is effectively invisible to the detector, leading to just the two top quarks and the missing energy from the $\tilde{\chi}_1^0$ being reconstructed.

Data and Monte Carlo Samples

This study was performed using 36.07 fb^{-1} of data recorded by the ATLAS detector during 2015 and 2016 data-taking. The data were recorded using the E_T^{miss} trigger as in the previous chapter - the threshold of the trigger depends on the instantaneous luminosity delivered by the LHC, but the

maximum threshold reached was 110 GeV. More details can be found in Section 6.2, as the same data was used for both studies.

Similarly to the data, the same background MC samples were used for both this reinterpretation study and the original stop decay analysis - details of the simulation can be found in Section 6.2. The main backgrounds are Z + jets processes, where the Z decays to neutrinos, single top production in association with a W boson where one W decays hadronically and the other decays leptonically, and $t\bar{t}$+V processes where the tops decay hadronically and the vector boson decays leptonically. Semileptonic $t\bar{t}$ production, W + jets and diboson backgrounds are all present but at a lower level than the three main background sources. QCD multijet and hadronic $t\bar{t}$ background contributions are small, and are estimated using the same jet smearing method described in Appendix B.

The gluino-mediated $\tilde{t}_1$ production signal is simulated using MadGraph5_AMC@NLO interfaced to Pythia 8 and EvtGen, similarly to the direct stop pair production samples. At each signal mass point, defined by the $\tilde{g}$ and $\tilde{\chi}_1^0$ masses, the mass splitting between $\tilde{t}_1$ and $\tilde{\chi}_1^0$ is set to $\Delta\, m(\tilde{t}_1, \tilde{\chi}_1^0) = 5$ GeV. Signal cross sections are calculated to next-to-leading order in the strong coupling constant, adding the resummation of soft gluon emission at next-to-leading-logarithmic accuracy. Again, more details can be found in Section 6.2.

Signal, Control and Validation Regions

After requiring the data to have passed the trigger requirement, and meeting the reconstructed primary vertex requirement, events must pass the same preselection requirements as for the nominal $\tilde{t}_1$ analysis. These requirements are shown in Sections 6.4 and 6.5, and are summarised in Table 7.1 below for convenience.

This reinterpretation was previously presented at the ICHEP 2016 conference, using 13.3 fb^{-1} of $\sqrt{s} = 13$ TeV data recorded in 2015 and 2016 by ATLAS [82]. The signal region definition for that analysis is shown in Table 7.2, with the results shown in Figure 7.4. No excess was seen, and the limits set managed to exclude the entire signal mass grid, in the $\tilde{g}$-$\tilde{\chi}_1^0$ plane - the numbers in Figure 7.4 are the cross-section limits for each mass point.

Given that the entire mass grid of simulated signal samples was excluded, an extension to the grid was produced and new samples simulated. The extended grid is shown in Figure 7.5 - the gluino mass reaches up to 2 TeV, with a $\tilde{\chi}_1^0$ mass ranging from 200 GeV to 1100 GeV at intervals of 100 GeV in both sparticle masses. The samples were generated in the same way as the previous grid, using MadGraph5_AMC@NLO, Pythia 8 and EvtGen.

Selection	**0 Lepton (SR)**	**1 Lepton (VR)**	**2 Lepton (CR)**
Trigger	E_T^{miss}	E_T^{miss}	Lepton p_T
N_{lep}	0	1	2
N_{jet}	≥ 4		
$N_{\text{b-tag}}$	≥ 2	≥ 1	
Jet p_T	$> 80, 80, 40, 40$ GeV		
E_T^{miss}	> 250 GeV		< 50 GeV
E_T^{miss},	-		> 100 GeV
$\lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert$	> 0.4		-
$m_T(l, E_T^{\text{miss}})$	-	< 100 GeV	-
m_{ll}	-		$86 < m_{ll} < 96$ GeV
$E_T^{\text{miss,track}}$	> 30 GeV	-	
$\lvert\Delta\phi(E_T^{\text{miss}}, E_T^{\text{miss,track}})\rvert$	$< \pi/3$	-	

Table 7.1: Preselection requirements for signal, control and validation regions.

Variable	**ICHEP SR**
Preselection	0 Lepton
b-tagged jets	≥ 2
$m^0_{\text{jet},R=0.8}$	> 120 GeV
$m^1_{\text{jet},R=0.8}$	> 60 GeV
$m_T^{b,\text{min}}$	> 175 GeV
E_T^{miss}	> 250 GeV
H_T	> 1100 GeV
$E_T^{\text{miss}}/\sqrt{H_T}$	$> 15\sqrt{\text{GeV}}$

Table 7.2: Selection criteria used for the gluino-mediated signal region, shown at the ICHEP 2016 conference. [82]

The signal region definition must be re-optimised for these new, higher-mass signal models. This reoptimisation is done with TMVA [66], using the cut-based optimisation procedure. No new variables were added into the signal region, and the preselection was left untouched. Two signal mass points are chosen as benchmark points, to be representative of the rest of the mass grid: $m_{\tilde{g}} = 1700$ GeV, $m_{\tilde{\chi}_1^0} = 400$ GeV and $m_{\tilde{g}} = 1600$ GeV, $m_{\tilde{\chi}_1^0} = 1000$ GeV. The two benchmark signal points were

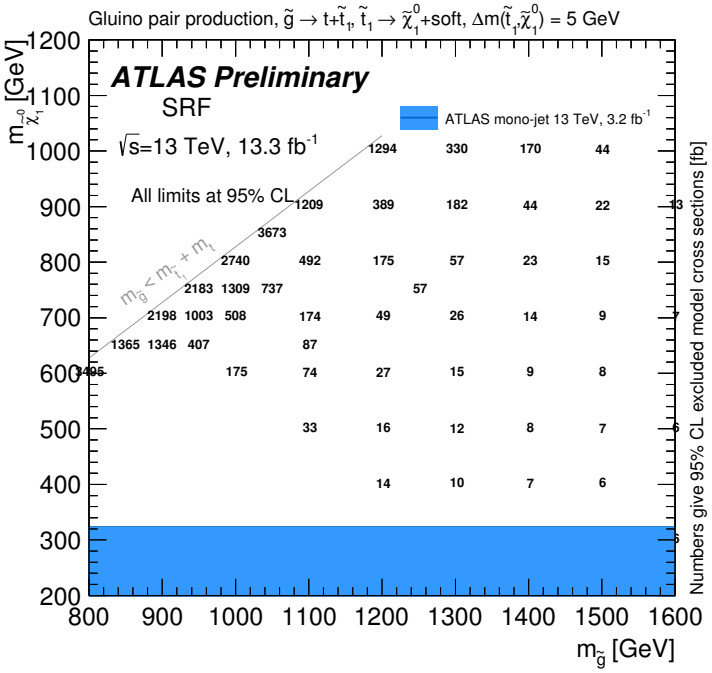

Figure 7.4: Cross section limits at 95% CL exlcusion as a function of $\tilde{g}$ and $\tilde{\chi}_1^0$ masses for gluino-mediated stop decays, $\tilde{g} \to t\tilde{t}_1 \to t\tilde{\chi}_1^0 +$ soft. This result was shown at the ICHEP 2016 conference, using 13.3 fb^{-1} of data. [82]

optimised individually, and the set of cuts (defined by their signal efficiency) that reaches the highest significance in both samples was used to define the new signal region. The significance is calculated as the ratio between the mean values of two Poisson distributions in a number-counting experiment [108, 109], assuming a 30% background uncertainty. Figure 7.6 shows the significance as a function of signal efficiency for the two benchmark points. The statistical fluctuations at low signal efficiency make it difficult to pick out a specific maximum significance, as a small step in either direction may cause a large change, hence the optimal signal efficiency was chosen to be 0.2, to be away from these low-stat regions. This signal efficiency corresponds to the cuts shown in Table 7.3, after taking into account the criteria favoured by both benchmark points.

The newly optimised signal region is described in Table 7.3 - the main differences between the new SR and the ICHEP SR are much tighter requirements on the E_T^{miss} and E_T^{miss}/H_T, while the H_T cut is much looser. The mass requirement on the sub-leading reconstructed W candidate, $m^1_{\text{jet},R=0.8}$, is also tighter. Specific control regions are defined for this signal region, for the Z + jets and semileptonic $t\bar{t}$ background processes (CRZE and CRTopE respectively), while for W + jets and single-top background sources the same control and validation regions from the direct stop pair production analysis, described

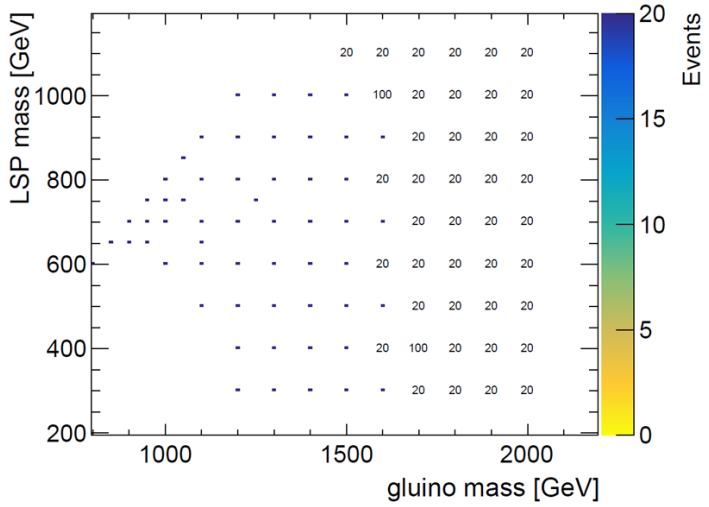

Figure 7.5: Extended grid for gluino-mediated stop decays. All grid points are 100 GeV apart in both $\tilde{g}$ mass and $\tilde{\chi}_1^0$ mass. The numbers at each point show the number of events generated in each sample - the two benchmark points have 100k events per sample, and the remainder of the grid have 20k events.

in Section 6.5, are used. The validation regions (VRZE and VRTopE) use an inverted cut on either $m^0_{\text{jet},R=0.8}$ or $m_T^{b,\text{min}}$ to stay orthogonal to the signal region, while loosening the cuts on other variables to increase the available statistics. This is particularly important for the VRTopE, where both the H_T and $E_T^{\text{miss}}/\sqrt{H_T}$ cuts have been removed, to reduce the level of signal contamination. Figure 7.7 shows the $N-1$ distributions of the discriminating variables in this newly optimised signal region, where all cuts except the one on the variable being plotted (the N in $N-1$) are applied, to demonstrate the effectiveness of the optimisation.

Figures 7.8 through 7.11 show distributions of a selection of discriminating variables in the specially-designed control and validation regions for this SR. The simultaneous fit across the CRs, as described in Section 6.5, has been performed to normalise the MC to the data, and includes the non-SR-specific control regions CRST, CRW and CRTTGamma.

Table 7.4 details the expected yields for the signal benchmark models and the background processes. The expected background yields are taken directly from MC; event yields post-fit can be found in the next section along with the unblinded signal region results. The QCD multijet and all-hadronic $t\bar{t}$ background processes are estimated using the jet-smearing method detailed in Appendix B.

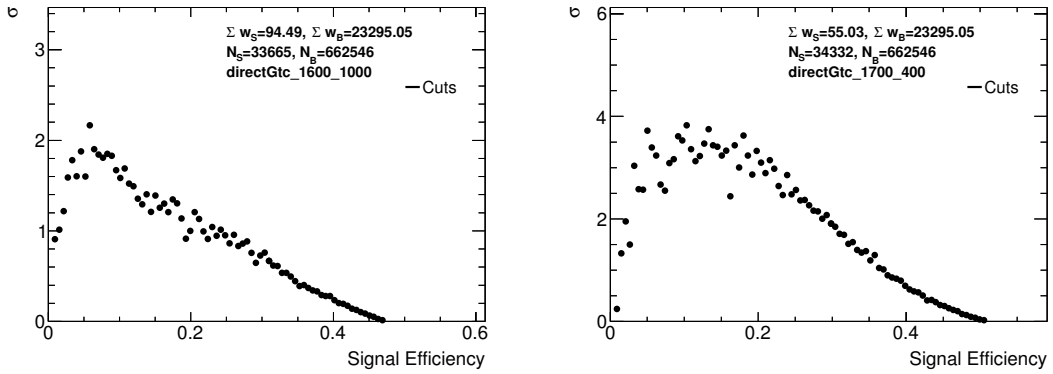

Figure 7.6: Significance vs signal efficiency for the cut-based optimisation, for *left*: $m(\tilde{g}, \tilde{\chi}_1^0) = (1600, 1000)$ GeV and *right*: $m(\tilde{g}, \tilde{\chi}_1^0) = (1700, 400)$ GeV. The chosen signal efficiency, to define the signal region cuts, was 0.2, to avoid the statistical fluctuations at low signal efficiency while still achieving a high significance.

	SRE	**CRZE**	**VRZE**	**CRTopE**	**VRTopE**
Preselection	0 Lepton	2 Lepton	0 Lepton	1 Lepton	0 Lepton
$N_{\text{b-tag}}$	≥ 2				
E_T^{miss}	> 550 GeV	< 50 GeV	> 250 GeV		
$E_T^{\text{miss}\prime}$	-	> 100 GeV	-		
$m_T^{b,\text{min}}$	> 200 GeV	-	> 200 GeV	> 100 GeV	$> 100, < 200$ GeV
$m_T^{b,\text{min}\prime}$	-	> 200 GeV	-		
$m^0_{\text{jet},R=0.8}$	> 120 GeV	-	< 120 GeV	> 120 GeV	
$m^1_{\text{jet},R=0.8}$	> 80 GeV	-		> 80 GeV	
H_T	> 800 GeV	> 500 GeV			-
$E_T^{\text{miss}}/\sqrt{H_T}$	$> 18\sqrt{\text{GeV}}$	-	$> 14\sqrt{\text{GeV}}$	-	-
$\Delta R(b,b)$	-		< 1 GeV	-	
$m_T(l, E_T^{\text{miss}})$	-			> 30 GeV	-
$\Delta R(b_{0,1}, l)_{\text{min}}$	-			< 1.5	-

Table 7.3: Selection criteria for the reoptimised gluino-mediated signal region, and the related Z + jets and $t\bar{t}$ control and validation regions. CRST, CRW and CRTTGamma, as defined in Chapter 6, are also included in the simultaneous fit for the background estimation.

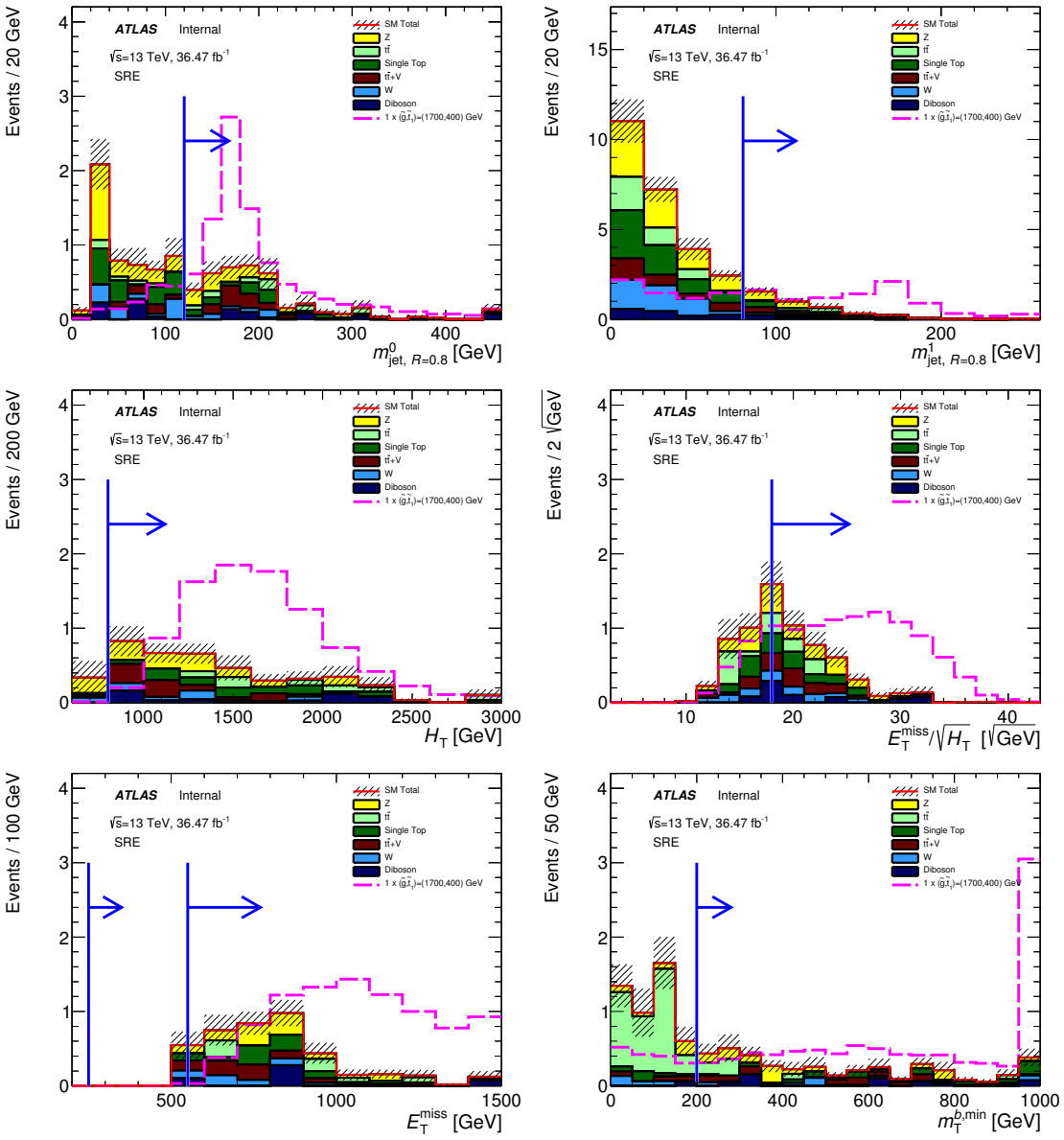

Figure 7.7: $N-1$ distributions for discriminating variables in the gluino-mediated stop decay signal region. All cuts except for the variable being shown are applied. *From top to bottom, left to right*: $m^0_{\text{jet},R=0.8}$, $m^1_{\text{jet},R=0.8}$, H_T, $E_T^{\text{miss}}/\sqrt{H_T}$, E_T^{miss}, and $m_T^{b,\text{min}}$. The MC is normalised to 36.07 fb^{-1} data. The signal shown as the pink dashed histogram is the $(m_{\tilde{g}}, m_{\tilde{t}_1})$=(1700,400) GeV benchmark point. The blue line indicates the cut to be applied on that variable.

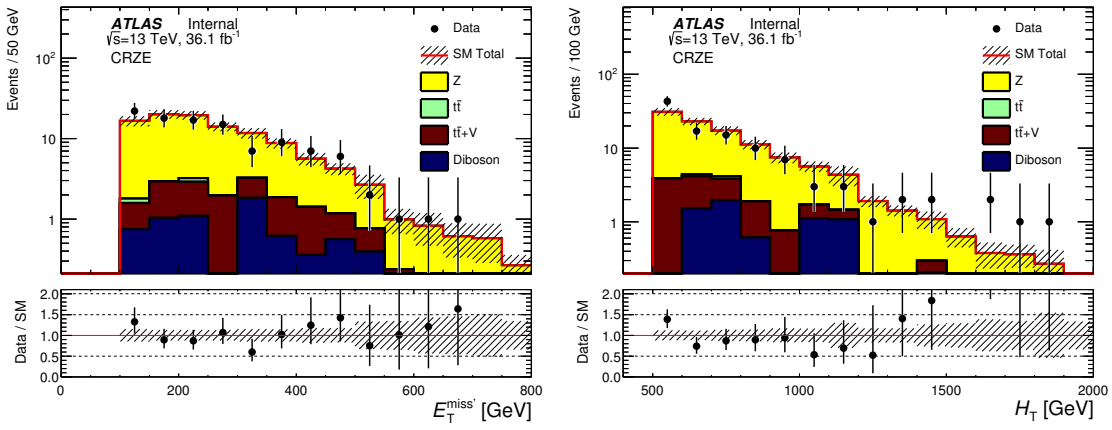

Figure 7.8: Distributions of $E_T^{\text{miss}\prime}$ (*left*) and H_T (*right*) for CRZE, for 36.07 fb^{-1} of data. The simultaneous fit to the CRs (including CRST, CRW and CRTTGamma as defined in Chapter 6) to normalise MC to data has been performed. The hatched area represents the uncertainty due to MC statistics and detector systematics.

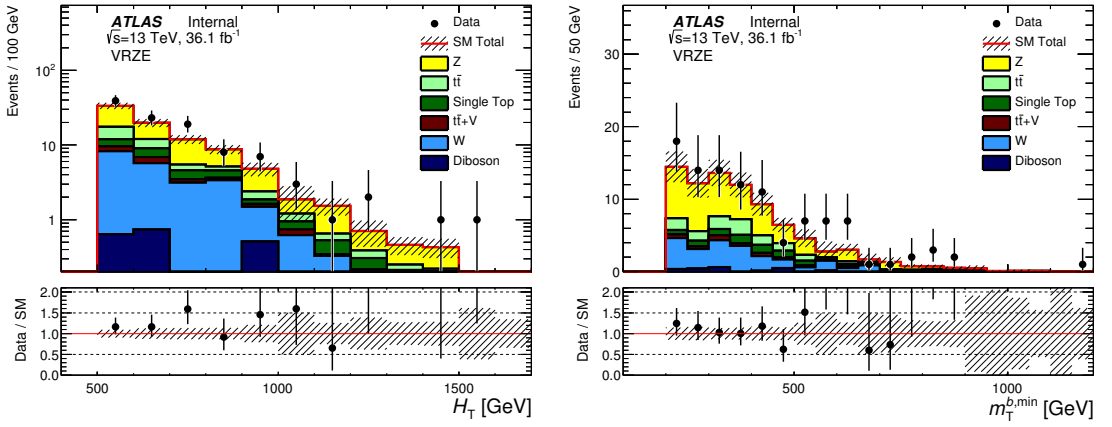

Figure 7.9: Distributions of H_T (*left*) and $m_T^{b,\min}$ (*right*) for VRZE, for 36.07 fb^{-1} of data. The simultaneous fit to the CRs (including CRST, CRW and CRTTGamma as defined in Chapter 6) to normalise MC to data has been performed. The hatched area represents the uncertainty due to MC statistics and detector systematics.

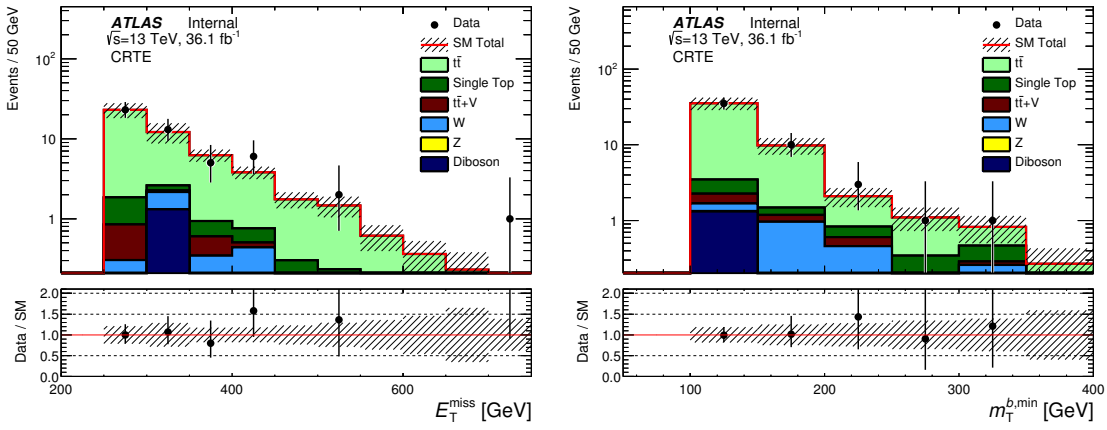

Figure 7.10: Distributions of E_T^{miss} (*left*) and $m_T^{b,\text{min}}$ (*right*) for CRTopE, for 36.07 fb^{-1} of data. The simultaneous fit to the CRs (including CRST, CRW and CRTTGamma as defined in Chapter 6) to normalise MC to data has been performed. The hatched area represents the uncertainty due to MC statistics and detector systematics.

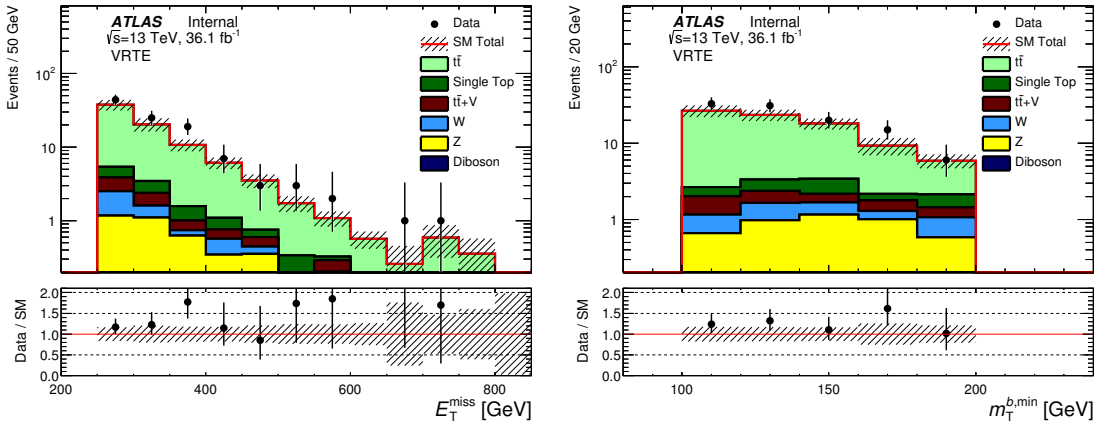

Figure 7.11: Distributions of E_T^{miss} (*left*) and $m_T^{b,\text{min}}$ (*right*) for VRTopE, for 36.07 fb^{-1} of data. The simultaneous fit to the CRs (including CRST, CRW and CRTTGamma as defined in Chapter 6) to normalise MC to data has been performed. The hatched area represents the uncertainty due to MC statistics and detector systematics.

SRE	
Z + jets	1.15 ± 0.18
Dibosons	0.00 ± 0.00
$t\bar{t}$	0.21 ± 0.06
Single top	0.56 ± 0.14
$t\bar{t}$ + Z	0.69 ± 0.13
W + jets	0.42 ± 0.13
Total MC	3.02 ± 0.30
($m_{\tilde{g}}$, $m_{\tilde{t}_1}$)=(1600,1000) GeV	5.93 ± 0.16 (2.1σ)
($m_{\tilde{g}}$, $m_{\tilde{t}_1}$)=(1700,400) GeV	9.33 ± 0.15 (3.1σ)

Table 7.4: Expected event yields taken from MC simulation for the background and signal processes. All values are normalised to 36.07 fb^{-1} luminosity, and errors are MC statistics only. The quoted significances are for optimisation only, and are calculated assuming a 30% background uncertainty.

Results

The same systematic uncertainties described in Section 6.6 are taken into account in this study, and are calculated using the same methods. This signal region is dominated by the statistical uncertainty on the background estimation. In terms of detector-related systematic uncertainties, the uncertainty on the calculation of the soft term of the E_T^{miss} is the largest contributor, providing 7% uncertainty to the SR background estimation. The jet energy scale (JES) and jet energy resolution (JER) are the next two largest contributors, giving a combined uncertainty of 5%. The b-tagging efficiency uncertainty contributes 3%, and the uncertainty associated with pileup correction contributes 2%. The same 3.2% uncertainty on the measurement of the integrated luminosity is applied to this SR as well.

Theoretical uncertainties are evaluated in the same method as in Section 6.6, as a variation in the transfer factor from CR to SR. Various sources of theoretical uncertainty are combined to provide a single systematic uncertainty value per background source. The uncertainty associated to Z + jets background sources is 37%, and for W + jets is 10%. For the $t\bar{t}$ and $t\bar{t}$ + Z backgrounds, these values are 15% and 55% respectively. For single top production, the associated uncertainty is 18%, with an additional 30% uncertainty applied to account for the effect of interference between single top and $t\bar{t}$ production.

Signal-related systematics are only taken into account during the limit-setting process. These arise mostly from detector and acceptance effects, and are calculated in the same way as for the background contributions. Each signal mass point has a different uncertainty, hence only the range of values is stated here. The JES and JER are in general the largest contributors, ranging from 3 to 23% and 6 to 16% respectively. The uncertainty due to pileup corrections ranges from 6 to 20%. Any uncertainty on the estimation of the number of signal events due to cross-section uncertainties is taken into account by calculating two additional limits, corresponding to a $\pm 1\sigma$ change in the predicted cross-section. The cross-section uncertainty itself is $\sim$ 20%, slightly larger for gluino-mediated stop production than for direct stop production.

The observed event yields, and the fitted number of background events, is shown in Table 7.5. The total background estimate is determined from the simultaneous fit to all control and validation regions, as was done for the direct stop pair production signal regions, described in the previous chapter. The errors on the MC yields include both statistical uncertainty and systematic uncertainty - the systematic uncertainties are the same as were included in the nominal $\tilde{t}_1$ decay analysis, described in Section 6.6. Figure 7.12 shows the discriminating variables with the MC overlaid with data and the background normalised to the yields extracted from the simultaneous fit.

No excess above the SM expectation is observed in the signal region, with a p-value of just 0.50 under the background-only hypothesis, so limits are set. The 95% CL upper limits on the number of signal

SRE	
Observed	3
Total SM	3.6 ± 1.11
$t\bar{t}$	$0.21^{+0.39}_{-0.21}$
W + jets	0.52 ± 0.27
Z + jets	1.36 ± 0.56
$t\bar{t}$+ V	0.89 ± 0.23
Single top	$0.66^{+0.70}_{-0.66}$
Dibosons	–
Multijets	–

Table 7.5: Observed event yields in the gluino-mediated stop decay signal region, and fitted background yields, using 36.07 fb^{-1} of data. Errors include MC statistical, detector-related systematics and theoretical uncertainties on the extrapolation from CR to SR.

events are derived using the CL$_s$ prescription [73] and calculated from asymptotic formulae [72]. A full description of the limit-setting procedure is found in Section 6.7 - though there is no need to statistically combine signal regions in this case as only this SR is sensitive to these decays, and it is not split into top reconstruction categories. Figure 7.13 shows the observed (red line) and expected (blue line) limits, interpreted in the $\tilde{g}$-$\tilde{\chi}_1^0$ mass plane. Gluino masses are excluded below 1800 GeV, for values of $m_{\tilde{\chi}_1^0}$ below 800 GeV.

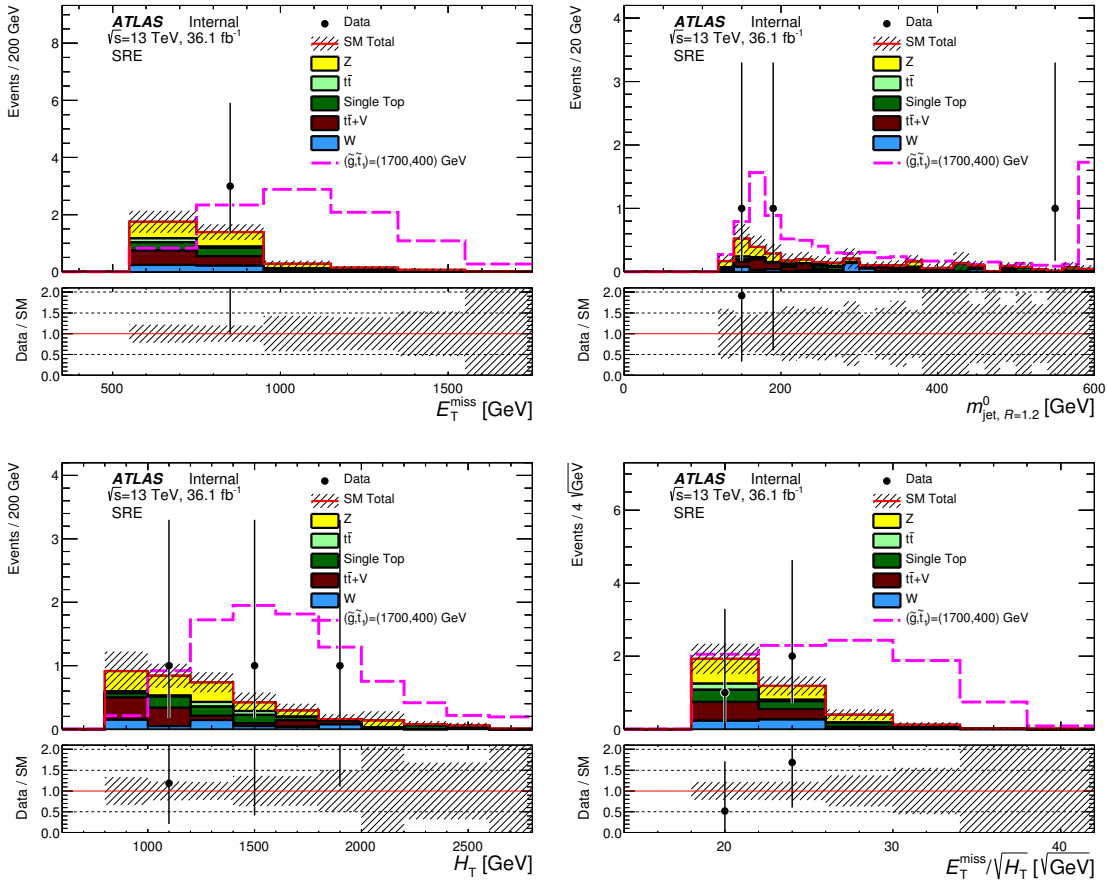

Figure 7.12: Distributions of discriminating variables in the gluino-mediated $\tilde{t}_1$ decay signal region, using 36.07 fb^{-1} of data. The MC background is normalised according to the simultaneous fit, and the hashed area indicates the uncertainty band, including statistical and systematic uncertainties. *From left to right, top to bottom*: E_T^{miss}, $m^0_{\text{jet},R=1.2}$, H_T and $E_T^{\text{miss}}/\sqrt{H_T}$.

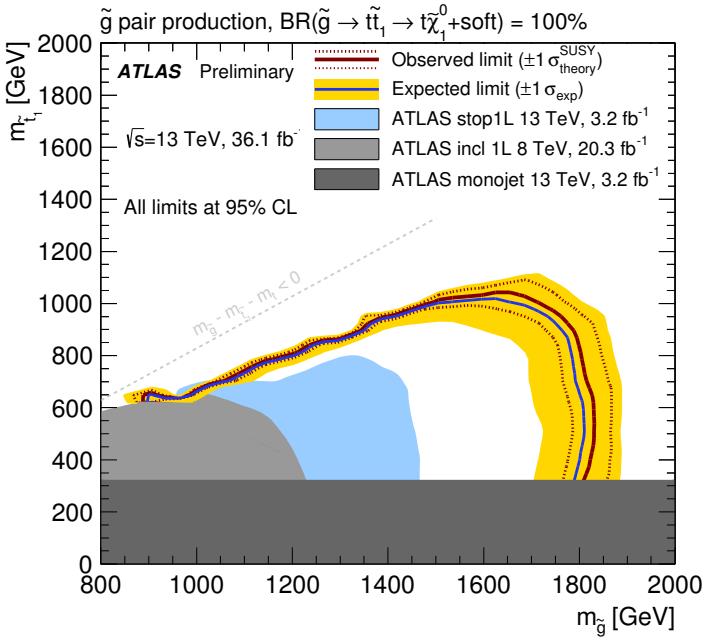

Figure 7.13: Expected (blue line) and observed (solid red line) exclusion limits at 95% CL as a function of $\tilde{g}$ and $\tilde{\chi}_1^0$ masses, for the scenario where $\tilde{t}_1$s are produced via $\tilde{g}$ decays, $\tilde{g} \rightarrow t\tilde{t}_1(\rightarrow \tilde{\chi}_1^0 + \text{soft})$. Uncertainty bands corresponding to $\pm 1\sigma$ variation on the expected limit (yellow band) and the sensitivity of the observed limit to $\pm 1\sigma$ variations of the signal theoretical uncertainties (red dotted lines) are also indicated. The numbers centered on the grid points show the expected CL_s values. The dark grey area is a limit on the mass of the $\tilde{\chi}_1^0$ placed by a previous ATLAS search for monojet processes that is independent of the gluino mass [110].

7.2 Direct Dark Matter Production

Supersymmetry provides an ideal candidate for dark matter (DM) [20, 79, 80] - the lightest supersymmetric partner particle, $\tilde{\chi}_1^0$, does not decay due to R-parity conservation, interacts via only the weak force, is massive and is electrically neutral; a so-called WIMP. But by searching for it, several assumptions are made that may not be accurate - first and foremost, that SUSY exists and conserves R-parity, but also about the hierarchy of sparticles and their interaction properties. If searching for direct dark matter production is the objective, a more model-independent approach would be ideal, where the only assumptions are the mass of the DM particle, and the spin of the mediator particle coupling it to the SM sector.

The ATLAS collaboration already has a number of dedicated dark matter searches, that generally fall into two categories: contact interaction searches and simplified model searches. [111–113]. Contact interactions use effective field theories to model dark matter production, where the model remains blind to the production mechanism, or the way in which the DM interacts with SM particles. In general the only final state particles in these models are the DM particles themselves, which are invisible to the detector - searches for such models often require the presence of initial state radiation, be it a high-p_T photon, jet or lepton, that can be used to 'tag' the event and boost the DM particle in the transverse plane. However, these models are usually only valid when the momentum transfer from mediator to DM particle is much smaller than the mass of the mediator, which is generally not true at high-energy colliders. The simplified model searches are more similar to SUSY searches - where a number of assumptions about the interactions between DM and SM particles are made - and this makes them useful tools to reinterpret existing data with existing analysis methods.

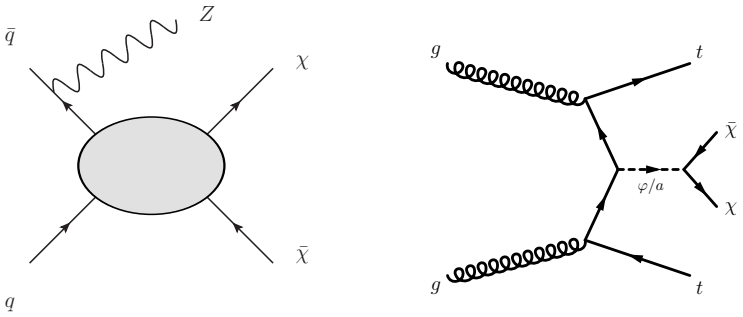

Figure 7.14: Examples of a contact interaction DM model (*left*) and a simplified DM production model (*right*).

It is possible to reinterpret the direct $\tilde{t}_1$ pair production analysis in terms of a DM simplified model, whereby a top quark pair is produced in association with a pair of DM particles (χ) produced via a

spin-0 mediator (ψ or a). This model is demonstrated in Figure 7.14 (*right*), and the mediator can be scalar or pseudo-scalar, coupling to the SM particles by mixing with the SM Higgs or extended Higgs sector [114]. The final state has two top quarks and large missing transverse momentum - the same final state as the nominal direct $\tilde{t}_1$ pair production analysis.

This DM reinterpretation formed part of the direct $\tilde{t}_1$ pair production analysis taken to ICHEP 2016 [82], with a dedicated signal region designed to target the unique elements of the DM production topology. Signal MC samples were generated using MadGraph5_AMC@NLO interfaced to Pythia 8 and EvtGen, with different values for both the $\psi(a)$ mass and the χ mass, under both scalar and pseudo-scalar mediator assumptions. The coupling of the mediator to the DM particles is assumed to be equal to its coupling to SM quarks, and is set to a value of 3.5, at the perturbative limit. The signal region that targets these decays is given in Table 7.6, and the observed and expected event yields are given in Table 7.7, for the 13.3 fb^{-1} of data used in the ICHEP publication. No excess was observed, and limits were set, shown in Figure 7.15, interpreted in the $m(\psi/a) - m(\chi)$ plane for both scalar and pseudo-scalar mediators. The same procedures for background estimation, systematic determination and limit-setting as described in Chapter 6 are used - see Ref. [82] for the precise systematic uncertainty values.

Variable	**DM SR**
b-tagged jets	$\geq$2
$m^0_{\text{jet},R=1.2}$	$>$ 140 GeV
$m^1_{\text{jet},R=1.2}$	$>$ 60 GeV
$m_T^{b,\text{min}}$	$>$ 200 GeV
$\tau-$veto	yes
$\Delta R(b,b)$	$>$ 1.5
E_T^{miss}	$>$ 300 GeV
$E_T^{\text{miss}}/\sqrt{H_T}$	$>14\sqrt{\text{GeV}}$

Table 7.6: Selection criteria for the Dark Matter production signal region, used in the ICHEP 2016 analysis. The same preselection criteria presented in Table 6.1 is also applied.

Given that the final state of these dark matter production events is so similar to the $\tilde{t}_1$ pair production final state - much more so than, for example, the gluino-mediated stop decay scenario - it is interesting to test the sensitivity of the nominal $\tilde{t}_1$ signal regions to the DM models. If the sensitivity is high enough, it may be possible to reinterpret the existing signal regions in terms of the new model, doubling the usage of the same data. This was tested using the ICHEP 2016 dataset with 13.3 fb^{-1} luminosity, with the two nominal $\tilde{t}_1$ production signal regions, SRA and SRB. The definitions used differ slightly

	DM SR
Observed	9
Total SM	7.1 ± 1.8
$t\bar{t}$	0.92 ± 0.48
W + jets	0.56 ± 0.17
Z + jets	2.78 ± 0.98
$t\bar{t} + V$	1.46 ± 0.55
Single top	$0.70 \, ^{+ 0.80}_{- 0.70}$
Dibosons	0.63 ± 0.48
Multijets	$0.01 \, ^{+ 0.02}_{- 0.01}$

Table 7.7: Observed and expected event yields for the DM SR, using 13.3 fb^{-1} of data. Background estimations are taken from MC simulation, after a simultaneous fit to all control and validation regions is used to normalise the simulation to data. Uncertainties include both statistical and systematic contributions.

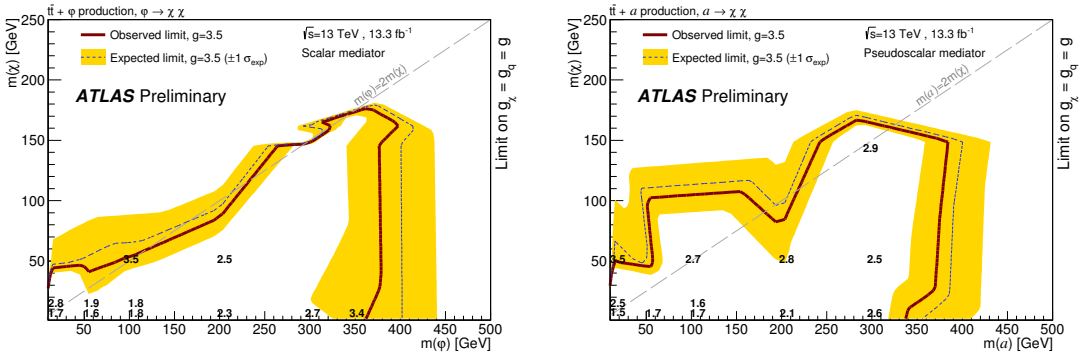

Figure 7.15: Expected (blue dashed line) and observed (red line) mediator coupling $g = 3.5$ exclusion limits at 95% CL as a function of ψ and χ masses (left) and a and χ masses (right) in the DM+$t\bar{t}$ scenario. Uncertainty bands corresponding to the $\pm 1\sigma$ variation on the expected limit (yellow band) are also indicated. The numbers on the plots indicate the limits of the coupling. Couplings above the perturbativity limit ($g = 3.5$) are not considered. [82]

to those used in the previous chapter, as the lesser amount of data requires looser cuts - plus the development of variables such as m_{T2} had not been as advanced at the time. Table 7.8 contains the definitions of the signal regions used to test sensitivity to DM models, and the results of this test are shown in Table 7.9, using the scalar DM model $m(\psi, \chi) = (350, 1)$ GeV as a benchmark model. The significance values shown are calculated assuming a 30% background uncertainty. None of the nominal

$\tilde{t}_1$ signal regions have a sensitivity higher than the dedicated DM SR defined above, which is perhaps unsurprising.

Signal Region		**TT**	**TW**	**T0**
	$m^0_{\text{jet},R=1.2}$	> 120 GeV	> 120 GeV	> 120 GeV
	$m^1_{\text{jet},R=1.2}$	> 120 GeV	$60 - 120$ GeV	< 60 GeV
SRA	$m^0_{\text{jet},R=0.8}$	> 60 GeV		
	b-tagged jets	≥ 2		
	$m_T^{b,\text{min}}$	> 200 GeV		
	τ-veto	yes		
	E_T^{miss}	> 400 GeV	> 450 GeV	> 500 GeV
SRB	b-tagged jets	≥ 2		
	$m_T^{b,\text{min}}$	> 200 GeV		
	$m_T^{b,\text{max}}$	> 200 GeV		
	τ-veto	yes		
	$\Delta R(b,b)$	> 1.2		
	E_T^{miss}	> 250 GeV		

Table 7.8: Selection criteria for SRA and SRB, used for the ICHEP 2016 publication and the DM sensitivity study. The preselection criteria from Table 6.1 are also applied.

Signal region	**DM signal events**	**SM Expectation**	**Significance**
SRA_TT	6.96 ± 0.15	5.25 ± 1.4	1.83
SRA_TW	8.36 ± 0.16	5.72 ± 1.6	2.06
SRA_T0	7.65 ± 0.15	11.5 ± 2.6	1.19
SRB_TT	12.2 ± 0.21	10.6 ± 2.3	1.97
SRB_TW	16.2 ± 0.24	16.8 ± 3.6	1.86
SRB_T0	21.9 ± 0.25	60.9 ± 14	0.81
DM SR	10.8 ± 0.17	7.14 ± 1.8	2.28

Table 7.9: Expected event yields for the $m(\psi, \chi) = (350, 1)$ GeV DM signal model, in the ICHEP 2016 versions of the $\tilde{t}_1$ production signal regions SRA and SRB, and the dedicated DM SR. Errors are from MC statistics only.

7.3 Searching for Compressed SUSY using MVA Techniques

As the region of phase-space where 'natural' high-mass SUSY can lie gets excluded further, more focus is placed on complex scenarios that may not be easily found. One such example is Compressed SUSY, where the mass difference between the lightest sparticles is small [115, 116]. This is challenging for traditional search styles that rely on missing energy in highly-boosted scenarios to separate the signal from the background - particularly in the case where the mass splitting between the $\tilde{t}_1$ and the $\tilde{\chi}_1^0$ is equal to the top mass. When $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0) = m_t$, the stop pair is produced roughly back-to-back, and the missing momentum from the two $\tilde{\chi}_1^0$s cancel out, resulting in a signature almost identical to $t\bar{t}$ production. Designing a signal region to separate these events is difficult, especially when only a cut-based approach is used.

Machine learning, or multivariate analysis (MVA), has been around for a long time, but its benefits to high-energy particle physics are just beginning to be explored. Recent papers [117–119] demonstrate the power of machine learning, both for excluding BSM models and for SUSY searches directly. The general principle is that an algorithm is allowed to 'learn' differences between signal and background events by being trained on a specific dataset with each event labelled as signal or background [120]. Different datasets can then be run through the algorithm, with each event being given a weight based on the output of the algorithm. If the training and the algorithm are effective, the signal and background weights will be pushed apart, often with very different distribution shapes. By then cutting at a certain value of the weight, or by fitting the shapes of the weight, signal and background events can be separated.

The compressed SUSY scenario discussed above was targeted during the $\tilde{t}_1$ pair production analysis, with the dedicated signal region SRC (see Section 6.4). This signal region uses the presence of an ISR jet to boost the stop pair and the $\tilde{\chi}_1^0$ in the transverse plane, then utilises a recursive jigsaw reconstruction algorithm to define the ISR and sparticle systems (see Appendix A for a full description of this method). Various kinematic quantities can then be constructed and used to separate the signal and background. However, simply cutting at a single value in each variable may not be able to fully exploit differences in the shape of the variables, especially in the cases where the variable shapes are not Gaussian. For this reason, this region of phase-space was chosen to test the capabilities of multivariate analysis techniques, as it stands to improve the most from a new approach that can efficiently separate the signal from the background.

This study is based upon the ICHEP 2016 $\tilde{t}_1$ pair production analysis, using 13.3 fb^{-1} luminosity [82]. The signal region designed for that luminosity differs to the one used for the full 36.07 fb^{-1} 2015 and 2016 dataset, and is summarised in Table 7.10. The main difference is that the bins in the R_{ISR} variable are overlapping, preventing any statistical combination - instead, the single bin with the

highest significance is the one chosen for exclusion purposes.

Variable	**SRC1**	**SRC2**	**SRC3**	**SRC4**	**SRC5**	**SRC6**	**SRC7**	**SRC8**
min R_{ISR}	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60
max R_{ISR}	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75
b-tagged jets	≥ 2				≥ 1			
N_{jet}^{S}	≥ 5							
p_{T}^{ISR}	> 400 GeV							
$p_{T,b}^{0,S}$	> 40 GeV							
$p_{T}^{4,S}$	> 50 GeV							
m_{T}^{S}	> 300 GeV							
$\Delta\phi_{\mathrm{ISR},I}$	> 3.0 radians							

Table 7.10: Selection criteria for the compressed SUSY scenario signal region, used in the ICHEP 2016 analysis with 13.3 fb^{-1} of data. The same preselection criteria from Table 6.1 are applied.

Choosing the variables to include in the MVA training is very important - the strength of the MVA heavily depends on the variables used. However, training on too many variables may actually reduce the effectiveness of the MVA if the input samples are not large enough, as the output becomes too narrowly focused, picking out specific details of the samples used rather than being generalised. This effect, known as overtraining, is well-studied and various methods of assessing the level to which an MVA output has been overtrained have been defined [121, 122]. For this study, to keep it simple and easily comparable with the nominal cut-and-count analysis, some of the variables used in the SR defined in Table 7.10 are included in the MVA training, with a few additions; the full list of variables is given in Table 7.11, with signal-background comparisons shown in Figures 7.16 and 7.17. Several other variables were studied that could be useful in an MVA but are unexploitable in a standard cut-and-count analysis, such as the $\Delta\phi(\mathrm{jet}^{i}, E_{T}^{\mathrm{miss}})_{[i=1,2,3]}$, where the shape of the signal and background differ but not enough for a single cut to distinguish them. However their effect on the output was found to be minimal, in this case - though their discriminating strength did appear to increase with $\tilde{t}_1$ mass, indicating potential future use. The linear correlations between the chosen variables in both the signal and background training samples are summarised in Figure 7.18 - uncorrelated variables are highly preferred, as then each variable provides a unique topology for the MVA to exploit.

Variable	**Definition**
N_{jets}	The number of jets in the event
E_T^{miss}	Missing transverse energy
$m_T^{b,\text{min}}$	Transverse mass between the b-jet closest in ϕ to the E_T^{miss} and the E_T^{miss}
$\min\ \lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert$	Minimum distance between either of the two leading-p_T jets and the E_T^{miss}
$\Delta\, R(b,b)$	Angular separation between the two jets with the highest b-tagging weights
R_{ISR}	Ratio between the E_T^{miss} and the p_T^{ISR}
m_T^S	Transverse mass between the whole sparticle system and the E_T^{miss}
$p_{T,b}^{0,S}$	p_T of the leading b-jet in the sparticle hemisphere
$p_{T,\text{jet}}^{3,S}$	p_T of the fourth highest-p_T jet in the sparticle hemisphere
N_{jets}^S	Jet multiplicity of the sparticle system
$N_{\text{b-tag}}^S$	Number of b-tagged jets in the sparticle system
$\lvert\Delta\phi(\text{ISR}, E_T^{\text{miss}})\rvert$	Angular separation between the ISR system and the E_T^{miss}

Table 7.11: Definitions of the variables used for the training of the MVAs.

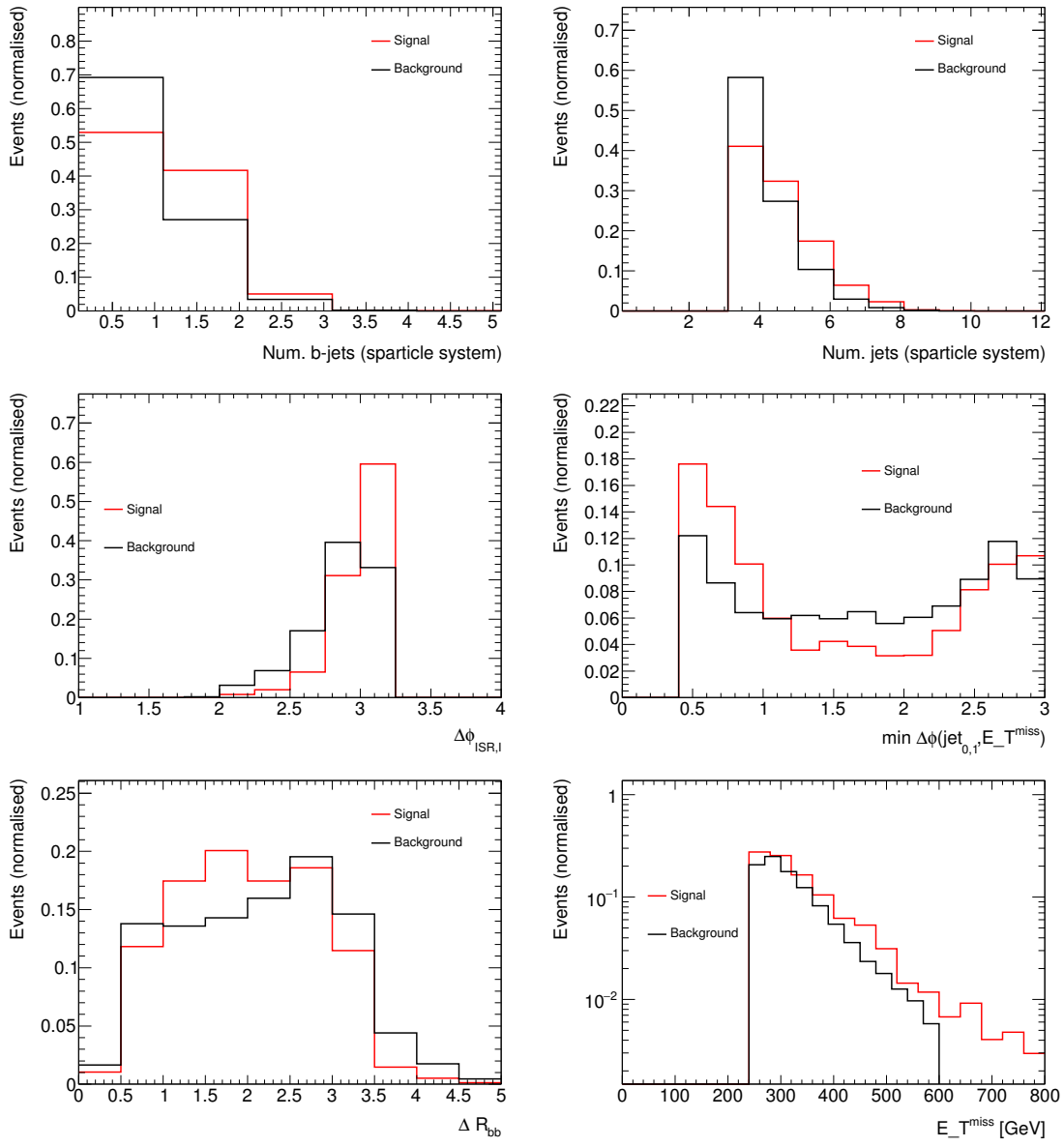

Figure 7.16: Background and signal distributions of (*from top to bottom, left to right*): $N^S_{\text{b-tag}}$, N^S_{jets}, $|\Delta\phi(\text{ISR}, E_T^{\text{miss}})|$, minimum $|\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})|$, $\Delta R(b,b)$ and E_T^{miss}. Histograms are normalised to unit area to compare shapes only. The signal used here is the (350,177) GeV sample.

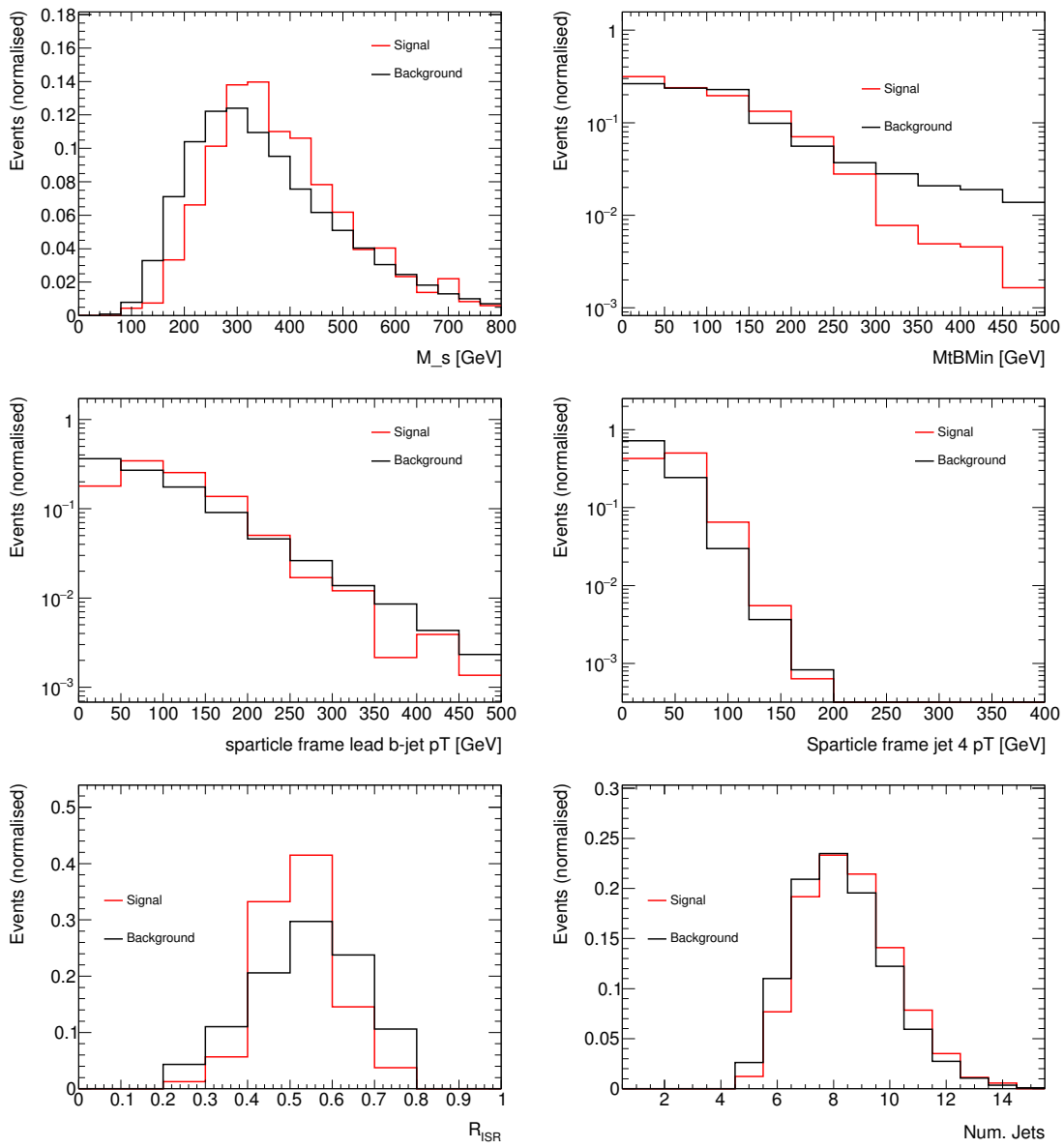

Figure 7.17: Background and signal distributions of (*from top to bottom, left to right*): m_T^s, $m_T^{b,\min}$, $p_{T,b}^{0,S}$, $p_{T,\text{jet}}^{3,S}$, R_{ISR}, and N_{jets}. Histograms are normalised to unit area to compare shapes only. The signal used here is the (350,177) GeV sample.

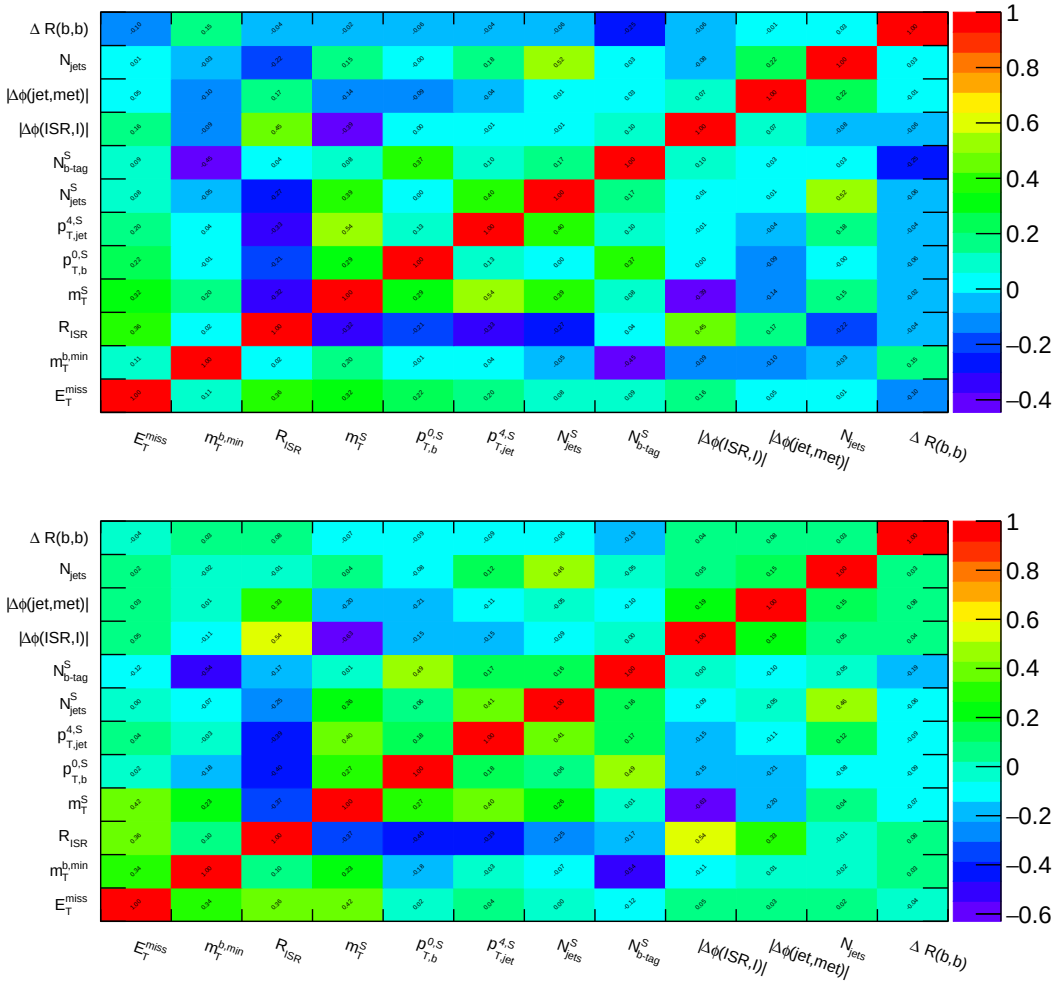

Figure 7.18: Linear correlation matrices for the training signal (*top*) and background (*bottom*) samples. In both cases, all included samples are combined. The preselection criteria from Table 7.12 is applied.

The MVAs were trained using TMVA [66], a ROOT-based machine learning software package that contains several different algorithms. Background samples included in the training sample were $t\bar{t}$, Z + jets, single-top and W + jets sources - the diboson and $t\bar{t} + Z$ background processes were found to have similar shapes to the other background sources, as well as have small contributions in the SR, and were thus removed from the training sample in order to speed up computing time. For the signal, in order to increase the available statistics and to keep the MVA more generalised to different mass points, several signal samples were combined: $m(\tilde{t}_1, \tilde{\chi}_1^0) = (350, 177)$ GeV, $m(\tilde{t}_1, \tilde{\chi}_1^0) = (400, 227)$ GeV and $m(\tilde{t}_1, \tilde{\chi}_1^0) = (450, 277)$ GeV. All samples sit on the $\Delta\, m(\tilde{t}_1, \tilde{\chi}_1^0) = m_t$ diagonal, and thus share similar kinematics. The effect of increasing $\tilde{t}_1$ and $\tilde{\chi}_1^0$ mass are shown in Figures 7.19 and 7.20 respectively, for several discriminating variables. This combination of signal samples is possible as the number of events in both background and signal samples are normalised during the training, based on the individual event weights, so that the background-to-signal ratio is 1:1 - the cross-section weighting of the signal events thus determines the relative contribution of each mass point to the overall signal training sample.

Multivariate analyses take many forms - even the cut-based analysis used in traditional SUSY searches is a multivariate analysis, albeit the most simple implementation possible. To find which would be the optimal method for this study, two of the most popular forms of MVAs were compared using the same training samples and variables: a neural network and a boosted decision tree. Neural networks [120, 123, 124] are one of the oldest forms of machine learning, that were initially thought of as too computationally intensive to be feasible. Technology eventually caught up, and the increased capability of computer processing combined with advances in the underlying mathematics, and the advent of so-called Deep Learning [125], has led to neural nets now being widely used for industrial applications. Input variables are passed to 'nodes', where events are given weights for each variable. These per-event variable weights are passed to the next 'layer' where they are combined in different ways and a new weight calculated, and the process is repeated until the final output layer is reached. Boosted decision trees (BDTs) [120, 126, 127] are an adaptation to the decision tree method; events travel down a 'tree' with diverging 'branches' that start and end at decision nodes or 'leaves', where events are weighted based on the value of a variable before being passed down the branch to the next leaf or another variable. In a Boosted Decision Tree, events are given additional weights at each leaf if it looks particularly signal-like or background-like, essentially pushing the two samples away from each other and widening the separation. Several different methods of boosting exist [127–129], but the Gradient boost method has been chosen, as it was shown to perform best; though care must be taken as it is more prone to overtraining than other boosting methods.

A Neural Network and a BDT are trained on the signal and background samples listed previously, using the preselection criteria listed in Table 7.12, which are an adaptation of the full signal region preselection from Table 6.1 and the SRC definition given in Table 7.10. The MVAs are trained and

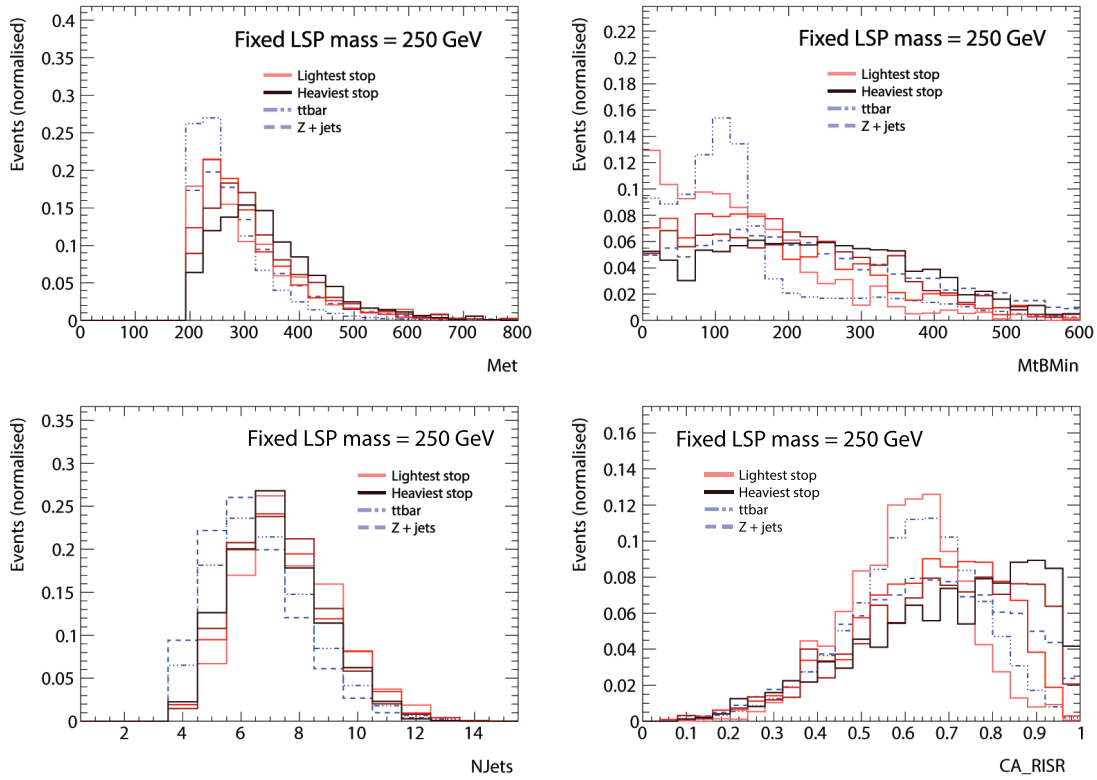

Figure 7.19: *From left to right, top to bottom*: distributions of E_T^{miss}, $m_T^{b,\text{min}}$, N_{jets} and R_{ISR} for MC signal samples with the $\tilde{\chi}_1^0$ mass fixed to 250 GeV and the $\tilde{t}_1$ mass ranging from 450 GeV to 600 GeV. The signal samples are shown as the solid histograms, with darker lines representing heavier $\tilde{t}_1$ masses, and $t\bar{t}$ and Z + jets backgrounds shown as the blue dashed lines. Histogram areas are normalised to unity in order to compare shape only.

tested using the Hold-out method [130], whereby the samples are split into two sub-samples, one for training and one for independently evaluating and testing the level of overtraining in the MVA.

Figure 7.21 shows the Receiver-Operator Characteristics (ROC) curves for the Neural Network, the BDT and a Fisher Discriminant (a linear discriminant that transforms variables into a space with zero linear correlations, shown for comparison only [120]) trained on the variables listed in Table 7.11. The ROC curves are an indicator of the ability of an MVA to separate the signal and background samples, showing the background rejection as a function of the signal efficiency. A perfect MVA, that can always pick out a signal event over a background event with no false positives, would have a background rejection of 1 for all values of signal efficiency, i.e. a horizontal line across the top of the signal efficiency-background rejection plane. In this case, none of the MVA methods shown achieve

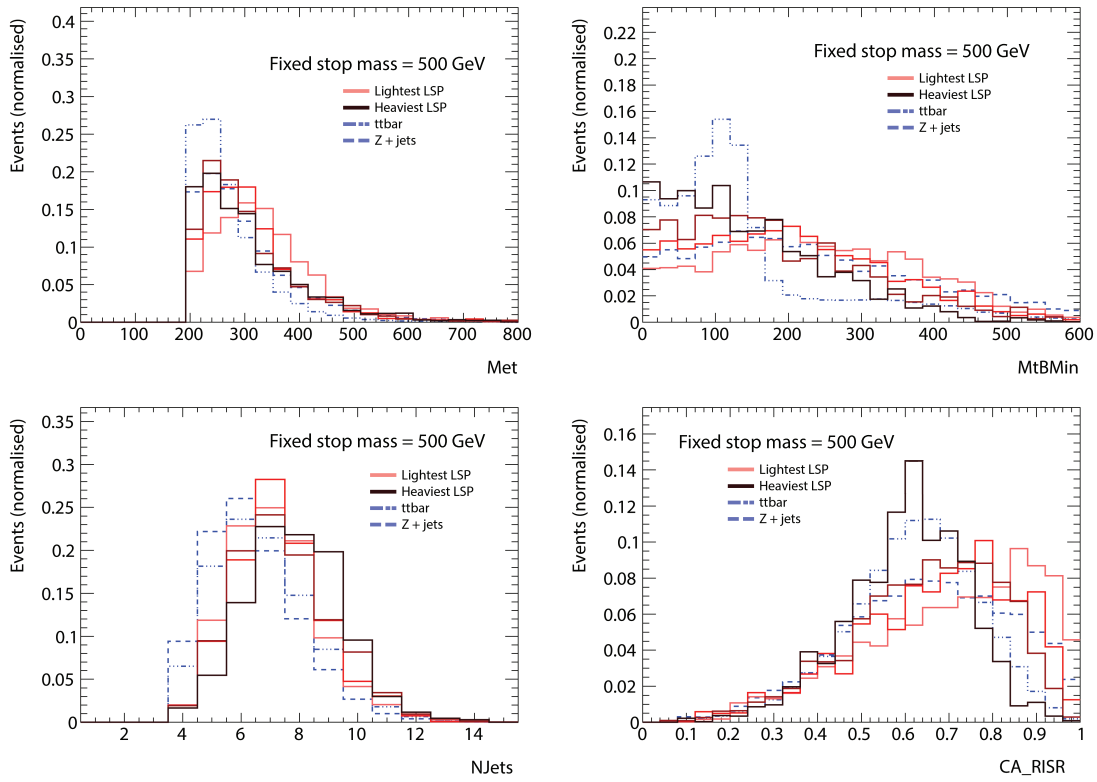

Figure 7.20: *From left to right, top to bottom*: distributions of E_T^{miss}, $m_T^{b,\text{min}}$, N_{jets} and R_{ISR} for MC signal samples with the $\tilde{t}_1$ mass fixed to 500 GeV and the $\tilde{\chi}_1^0$ mass ranging from 100 GeV to 327 GeV. The signal samples are shown as the solid histograms, with darker lines representing heavier $\tilde{\chi}_1^0$ masses, and $t\bar{t}$ and Z + jets backgrounds shown as the blue dashed lines. Histogram areas are normalised to unity in order to compare shape only.

perfect separation, which is to be expected given the complexity of this kind of $\tilde{t}_1$ decay.

As the ROC curves show, the BDT seems to do better than the Neural Net at separating the signal and background. This is perhaps unsurprising, given that Neural Nets require a large amount of data to reach an optimal configuration during training [122], whereas BDTs do better with lower statistics, as is the case here. However this could also be due to overtraining - if the BDT has reached an overtrained configuration, the ROC curve may show a false improvement. This can be checked by comparing the training sample to the testing sample - if the shapes of the MVA output in the two samples are vastly different, overtraining has occurred, as a small change in input values causes a large change in output value. A quantitative test of overtraining is obtained by applying a goodness-of-fit test between the two distributions. In this instance, the Kolmogorov-Smirnoff (KS) test is applied to the training

	Selection
N_{jets}	≥ 4
E_T^{miss}	> 250 GeV
Jet p_T	$>$ (80,80,40,40) GeV
$E_T^{\text{miss,track}}$	> 30 GeV
$\lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert_{\text{min}}$	> 0.4 radians
$N_{\text{b-tag}}$	≥ 1
p_T^{ISR}	> 400 GeV
R_{ISR}	$0.2 < R_{\text{ISR}} < 0.8$
N_{jets}^S	≥ 4

Table 7.12: Preselection criteria used to select events for the MVA training.

and testing samples. This goodness-of-fit test compares the distance between each point in each bin of the two histograms, and returns a probability that the samples come from the same underlying distribution. In theory, the output of the KS test is uniformly distributed between 0 and 1; if the output of the KS test is close to 1, the samples originate from the same distribution, with an output of 0 (or less than some critical value α, typically $\alpha < 0.05$) indicating the samples are incompatible. However, in practice, when comparing distributions once - as we are doing in this case, comparing the training and test samples from a single BDT - interpreting the KS test, as implemented within TMVA, becomes more similar to a χ^2 test; if the distributions are the same, what is the probability a lower value will be found? As a result, this effectively flips the interpretation of the KS test output, meaning a KS test value closer to 0 indicates the distributions come from the same underlying distribution. This comparison is shown in Figure 7.22 for both the Neural Net and the BDT - neither appear to be overtrained, indicating the BDT is the best choice for this study.

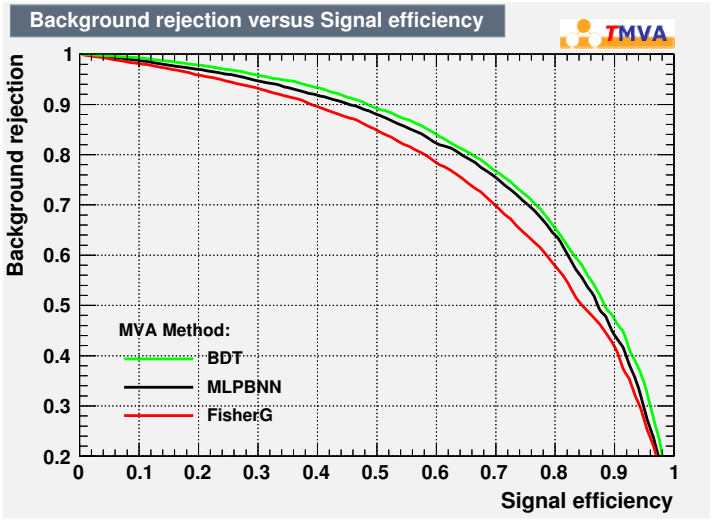

Figure 7.21: Receiver Operator Characteristics curves for the Neural Network (*black*), BDT (*green*) and Fisher Discriminant (*red*) MVA methods. The BDT is clearly superior to the Neural Net, in this particular use-case.

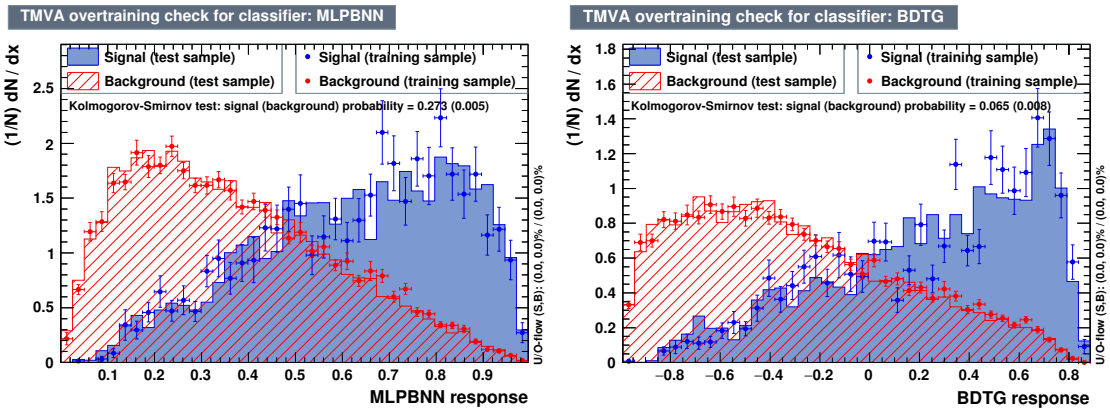

Figure 7.22: MVA output distributions for the training and test samples, for the Neural Net (*left*) and BDT (*right*) algorithms. Signal samples are shown in blue, with the training sample as the markers and the test sample as the shaded histograms, and the background samples in red. The Kolmogorov-Smirnov test statistic is shown on the plots, demonstrating that neither MVA is over-trained. Background samples are normalised to the signal histogram areas.

Results

Using the BDT output from Figure 7.22, the expected significance was calculated in steps of 0.1 in the BDT output for each of the signal samples. The value of the BDT that gave the highest significance across all three signal points was chosen as the optimal cut, which was found to be at BDT value > 0.69. This gives a signal efficiency of 16%, with respect to the events passing the training preselection from Table 7.12 alone, with a background rejection factor of 78. It also leaves signal statistics comparable to the highest-statistics bin of R_{ISR} in the cut-based SRC. Table 7.13 gives the expected number of events for signal and background, taken from simulation directly, along with the expected significance for each signal point used during the training stage for both the MVA and the cut-based analysis.

	SRC5	**SRC6**	**SRC-BDT**
Z + jets	$0.27^{+0.52}_{-0.27}$	$0.23^{+0.43}_{-0.23}$	0.83 ± 0.16
Dibosons	0.16 ± 0.13	0.16 ± 0.11	0.30 ± 0.16
$t\bar{t}$	9.7 ± 3.7	6.8 ± 3.5	14.41 ± 0.77
Single top	$0.54^{+0.64}_{-0.54}$	$0.48^{+0.56}_{-0.48}$	0.89 ± 0.16
$t\bar{t}$ + Z	0.26 ± 0.06	0.16 ± 0.11	0.33 ± 0.06
W + jets	0.68 ± 0.40	0.68 ± 0.23	0.91 ± 0.14
Total MC	11.6 ± 3.6	8.6 ± 3.5	17.67 ± 0.83
$(m_{\tilde{t}_1}, m_{\tilde{\chi}^0_1})$ = (350,177) GeV	21.57 ± 1.97 (4.2σ)	16.50 ± 1.75 (3.9σ)	34.14 ± 2.46 (4.5σ)
$(m_{\tilde{t}_1}, m_{\tilde{\chi}^0_1})$ = (400,227) GeV	15.43 ± 1.49 (2.9σ)	11.81 ± 1.69 (2.7σ)	22.36 ± 1.73 (3.1σ)
$(m_{\tilde{t}_1}, m_{\tilde{\chi}^0_1})$ = (450,277) GeV	8.14 ± 1.43 (1.8σ)	10.15 ± 1.64 (2.6σ)	15.39 ± 1.47 (2.3σ)

Table 7.13: Expected event yields for the BDT signal region for 13.3 fb^{-1} of data, applying the preselection cuts from Table 7.12 plus a BDT > 0.69 cut. Significances quoted are calculated assuming a 20% background uncertainty. The most significant SRs for each signal mass point from the original analysis are shown for comparison.

The highest significance achieved is for the $m(\tilde{t}_1, \tilde{\chi}^0_1) = (350, 177)$ GeV signal point, at 4.5σ - this compares favourably to the highest significance achieved in the cut-based analysis for the same signal point, at 4.2σ. This improvement is also found in the (400,227) GeV signal, which improves from 2.9σ to 3.1 sigma significance - however, the significance does decrease for the (450,277) GeV signal, going from 2.6σ to 2.3σ. This decrease is likely due to the relative abundance of each signal point in the training sample - the (450,277) GeV signal is weighted lower than the (350,177) GeV signal, meaning the BDT was trained on more of the lower-mass signal than the higher-mass signal, and the output and separation power represents that. This is compared to the cut-based analysis, which is binned in R_{ISR} to be sensitive to different signal points and be able to optimise individually for each.

If more statistics were available, a similar approach may be viable for the BDT - instead of bins in R_{ISR}, however, several BDTs may be trained separately, each targeting different clusters of mass points - for example, a low-mass BDT might target the points (300,127), (350,177) and (400,227), while a high-mass BDT might target (450,277), (500,327) and (550,377). Proceeding in this manner would eliminate the need for a single BDT to cover too many mass points with vastly differing topologies, that would cause the overall sensitivity to decrease.

Figure 7.23 shows a comparison of signal shapes in several discriminating variables, before and after the BDT cut is applied. This demonstrates the type of events the BDT is picking out, and how it decides whether an event looks signal-like or background-like. The shape of the distributions does not change massively, which is a good sign - it shows that no single variable is particularly important, but rather a combination of all variables is providing the separating power. Figure 7.24 shows the distributions of some of the discriminating variables post-BDT cut, with the signal overlaid on the background, to show the separation achieved by using the BDT. In most cases, the signal is well-separated, which lends itself well to interpreting data in terms of a signal model, if the SR was to be unblinded and an excess seen.

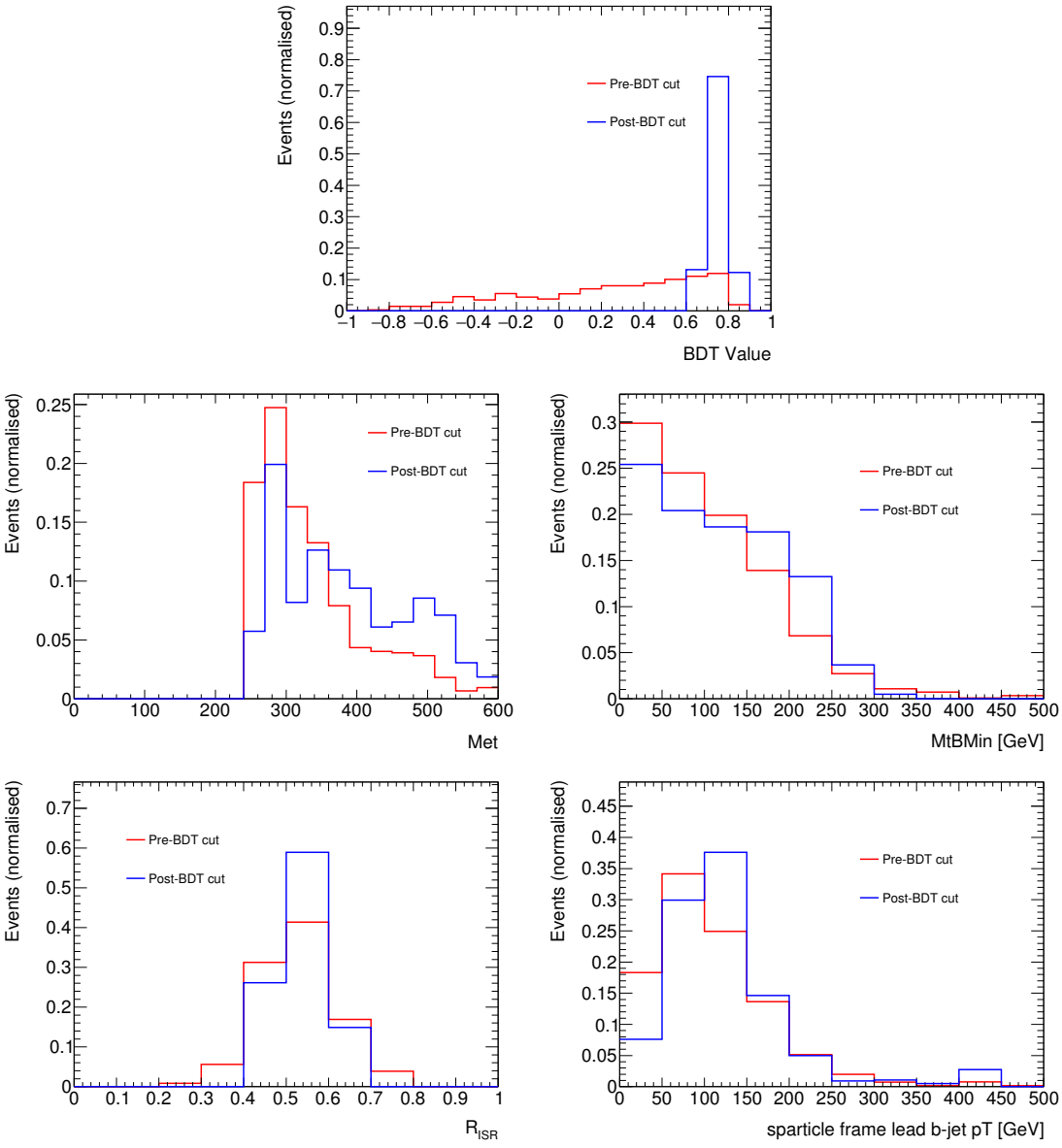

Figure 7.23: Distributions of discriminating variables for the $m(\tilde{t}_1, \tilde{\chi}_1^0) = (350, 177)$ GeV signal point, before (*red*) and after (*blue*) the BDT > 0.69 selection is applied. *From left to right, top to bottom*: BDT output, E_T^{miss}, $m_T^{b,\text{min}}$, R_{ISR}, $p_{T,b}^{0,S}$. Histogram areas are normalised to unity in order to compare shape only.

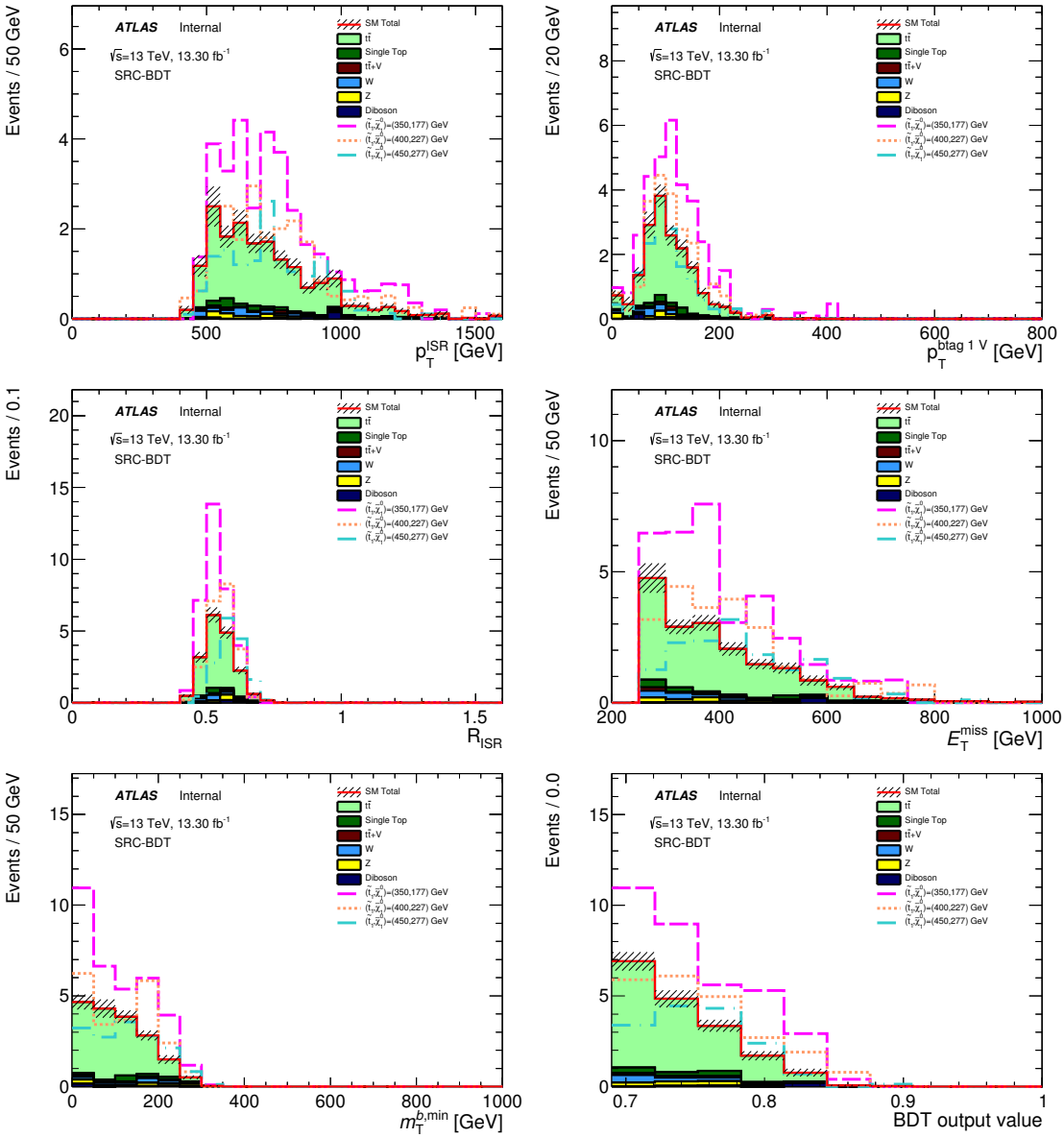

Figure 7.24: Distributions of several discriminating variables, and the BDT output itself, after the BDT > 0.69 cut is applied. *From left to right, top to bottom*: p_T^{ISR}, $p_{T,b}^{0,S}$, R_{ISR}, E_T^{miss}, $m_T^{b,\text{min}}$, and BDT output. Distributions are normalised to 13.3 fb^{-1} of data. The SM background expectation is shown as a stacked histogram, while the signal models are shown as the dashed lines. The hashed area on the SM background is from MC statistics only.

7.4 Summary

This chapter has presented the results from two reinterpretations of the existing all-hadronic $\tilde{t}_1$ pair production search, detailed in Chapter 6. The first reinterpretation, a gluino-mediated scenario in which gluinos are pair-produced and both decay into stops, with a small mass splitting between the $\tilde{t}_1$ and the $\tilde{\chi}_1^0$, is sensitive to gluino masses up to 2 TeV. After optimising the signal region for these high masses, the signal region was unblinded. No excess above the SM background expectation was seen, and limits on the $\tilde{g}$ and $\tilde{\chi}_1^0$ masses were set, excluding up to $m_{\tilde{g}} = 1800$ GeV for $\tilde{\chi}_1^0$ masses up to 800 GeV. This is comparable to the equivalent dedicated ATLAS gluino search [131] that excludes up to $m_{\tilde{g}} = 2.03$ TeV, but assumes a massless $\tilde{\chi}_1^0$. The second reinterpretation, a dark matter interaction scenario using simplified models to simulate the interaction between the dark matter and SM particles, was previously shown at the ICHEP 2016 conference [82], excluding dark matter mediator masses up to 370 GeV and 360 GeV, for scalar and pseudo-scalar dark matter, respectively. A sensitivity study was performed, to evaluate the sensitivity of the nominal $\tilde{t}_1$ decay signal regions to these dark matter production events - the maximum significance reached was 2.06σ, lower than the 2.28σ reached by the dedicated dark matter SR, and much lower than the significance expected for $\tilde{t}_1$ decays.

A study into the feasibility of using a multivariate analysis technique has also been presented in this chapter. A BDT was trained on simulated signal and background samples, using a combination of several signal mass points to both increase the available statistics and to generalise the BDT. The BDT was chosen from several different MVA methods, as it was the best performing in this particular scenario. An optimal cut on the BDT output was found, to maximise significance to the benchmark signal models, and the variables used to train the BDT were studied both before and after this cut is applied. Finally, the expected significance is calculated and compared to the nominal SR. The BDT was found to improve on the nominal cut-based signal region for two of the signal models, but the significance worsens for the highest-mass signal point due to the nature of the BDT training.

The interpretations detailed here are not the only options for extending the use of LHC data. Many more models exist that can make use of the $\tilde{t}_1$ pair production topology, both within SUSY and alternative models. For example, Stealth Supersymmetry [132] is an interesting version of SUSY that preserves R-parity but suppresses missing energy signatures by having light particles with small mass splittings. The phenomenological MSSM (pMSSM) [26, 133] is an adaptation of the MSSM that parameterises large numbers of the free MSSM variables, to create more complex signals than the simplified models widely used, but with well-defined signatures. ATLAS has previously set limits in the pMSSM parameter space using the LHC Run 1 data [134], and this may be done again when the LHC Run 2 is complete. There is still a lot of room to find some form of SUSY at the LHC, even if its more complex than originally thought.

Chapter 8

Conclusions

The switch-on of the LHC brought with it a plethora of world-class results from ATLAS, on topics ranging from the discovery of the Higgs boson to measurements of Standard Model processes like $B_s^0 \to \mu^+\mu^-$, to probing ever deeper into the phase-space of supersymmetry. While there have been tantalising glimpses of possible new physics, such as the now-infamous 750 GeV diphoton resonance [135], no excess has persisted through multiple datasets.

A measurement of the $B_s^0 \to \mu^+\mu^-$ branching ratio, performed with 25 fb^{-1} of data recorded by the ATLAS detector during $\sqrt{s} = 7$ and 8 TeV data-taking in 2011 and 2012, has been presented in Chapter 5. This improves on the previous ATLAS measurement that used just 4.9 fb^{-1} of data, detailed in Ref. [65], by including a BDT to remove fake-muon background sources and by optimising the BDT used to reduce continuum background sources. A small number of $B_s^0 \to \mu^+\mu^-$ events are observed, fewer than were expected, which is most likely due to an unfortunate statistical under-fluctuation. With the observed yield of $B_d^0 \to \mu^+\mu^-$ events constrained to be non-negative, a branching ratio for $B_s^0 \to \mu^+\mu^-$ is extracted with a value $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (0.9^{+1.1}_{-0.8}) \times 10^{-9}$, compatible with the Standard Model at the 2σ level. Allowing for the B_d^0 yield to be negative, limits are set on the B_s^0 branching ratio, excluding values above $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 3.0 \times 10^{-9}$ at 95% CL, which is also still compatible with the SM.

Also presented in this thesis is the search for all-hadronic direct $\tilde{t}_1$ pair production, using 36.07 fb^{-1} of $\sqrt{s} = 13$ TeV data recorded by ATLAS in 2015 and 2016. Supersymmetric top quarks, stops, are assumed to decay via either $\tilde{t}_1 \to t\tilde{\chi}_1^0$ or $\tilde{t}_1 \to b\tilde{\chi}_1^\pm \to bW^{(*)}\tilde{\chi}_1^0$, which provides a final state signature containing a high jet multiplicity, no leptons, and large E_T^{miss}. Several signal regions are defined to fully capture the unique topologies of all possible combinations of $m_{\tilde{t}_1}$ and $m_{\tilde{\chi}_1^0}$. No excess above the SM expectation is observed in any signal region, however, and limits are placed. Stop masses below 950 TeV are excluded, for models with $m_{\tilde{\chi}_1^0} < 160$ GeV, in the $\tilde{t}_1 \to t\tilde{\chi}_1^0$ channel, while for the

$\tilde{t}_1 \to b\tilde{\chi}_1^{\pm} \to bW^{(*)}\tilde{\chi}_1^0$ channel, $m_{\tilde{t}_1}$ is excluded below 860 GeV for $m_{\tilde{\chi}_1^0}$ values up to 210 GeV.

The data used for the $\tilde{t}_1$ pair production search has been reinterpreted in terms of a gluino-mediated stop production scenario, where gluinos decay via $\tilde{g} \to t\tilde{t}_1(\to q\tilde{\chi}_1^0)$. A specially-designed signal region is optimised to take advantage of the highly-boosted states this scenario produces, with both large E_T^{miss} and H_T. No excess above the SM expectation is observed, however, and limits are set on this process too, excluding gluino masses below 1800 GeV for $m_{\tilde{\chi}_1^0}$ values below 800 GeV. A brief look at another potential reinterpretation, a direct dark matter production scenario, is also presented in this thesis, using just 13.3 fb^{-1} of data recorded in 2015 and 2016. Limits were placed on both scalar and pseudo-scalar mediator masses [82].

Finally, a study into the potential of machine learning to enhance traditional SUSY search techniques is performed. Operating under the assumption of 13.3 fb^{-1} of data, in order to compare to the hadronic $\tilde{t}_1$ decay ICHEP 2016 result [82], a signal region is defined to look for compressed SUSY decays. A Boosted Decision Tree is found to perform best, and is trained on several signal models and background samples. A maximum significance of 4.5σ is achieved for the benchmark signal model, which does improve on the traditional cut-and-count analysis. However, some additional work would be required to fully understand the effects of the BDT on the signal, and to be sensitive to a range of different signal models, before this technique could be adapted into a full analysis.

In general, these results - and others like them from ATLAS, CMS and LHCb - paint a bleak picture in the continuing search for New Physics. Many datasets taken over many years, with different energies and different detectors, all say the same thing: we haven't found it yet. But as even more data starts to roll in, with the LHC due to run through 2017 and 2018, there is still room to manoeuvre. New, testable theories are emerging, and even within the framework of natural SUSY, new decay chains and models can be designed to fill in the gaps that haven't been excluded yet. In other words: there's no stopping now!

Appendix A

Recursive Jigsaw Reconstruction

Models with a small difference between the $\tilde{t}_1$ and $\tilde{\chi}_1^0$ mass, $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0) = m_t$, are experimentally difficult to search for. As the mass splitting approaches the top quark mass, the stop decay has little additional energy to give momenta to the $\tilde{\chi}_1^0$ and top systems, resulting in softer final state objects and E_T^{miss}. While the two stops can be strongly boosted back-to-back against one another in the di-stop rest frame, the $\tilde{\chi}_1^0$ and tops tend to be produced roughly at rest in their respective stop production frames, causing them to be approximately collinear to the original stop boost axis. This means the two $\tilde{\chi}_1^0$s will also have close to equal and opposite momenta in the di-stop frame, resulting in the combined di-$\tilde{\chi}_1^0$ system having a small vectorial p_T (and ultimately E_T^{miss}) unless the entire di-stop system is itself boosted in the transverse plane.

Traditional searches for R-parity conserving SUSY models rely on the presence of large-E_T^{miss} final states caused by the presence of two heavy $\tilde{\chi}_1^0$s. However this is not effective in the compressed regime, as cutting hard on E_T^{miss} alone will decrease the signal efficiency exponentially while the resulting top quarks from the stop decays will make the event seem like SM $t\bar{t}$ production.

The fact that the $\tilde{\chi}_1^0$ gains very little momenta from the stop decays leads to other useful correlations that can be exploited. If the di-$\tilde{\chi}_1^0$ system recoils against a system of ISR jets, it gets boosted in the transverse plane, resulting in a net missing energy contribution which will have a strong correlation with the momenta of the ISR system. To zeroth order, the correlation between E_T^{miss} and ISR p_T is proportional to the stop and LSP masses:

$$\frac{E_T^{\text{miss}}}{p_T^{\text{ISR}}} \sim \frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{t}_1}}. \tag{A.1}$$

This correlation is strongest when $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0) = m_t$, as shown in Figure A.1 for a 400 GeV stop and 227 GeV $\tilde{\chi}_1^0$ sample with fully hadronic decays.The slope of the red line through the middle of the

distribution is exactly the relationship shown in Eq. A.1 - the width of the distribution is related to the combination of top and stop widths.

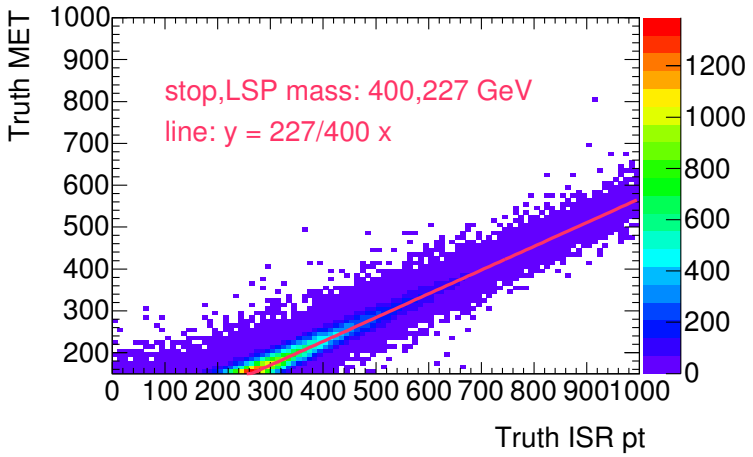

Figure A.1: Distribution of E_T^{miss} vs total ISR p_T at truth level for a $m_{\tilde{t}_1} = 400$ GeV, $m_{\tilde{\chi}_1^0} = 227$ GeV signal point. Only fully hadronic decays are included in the plot.

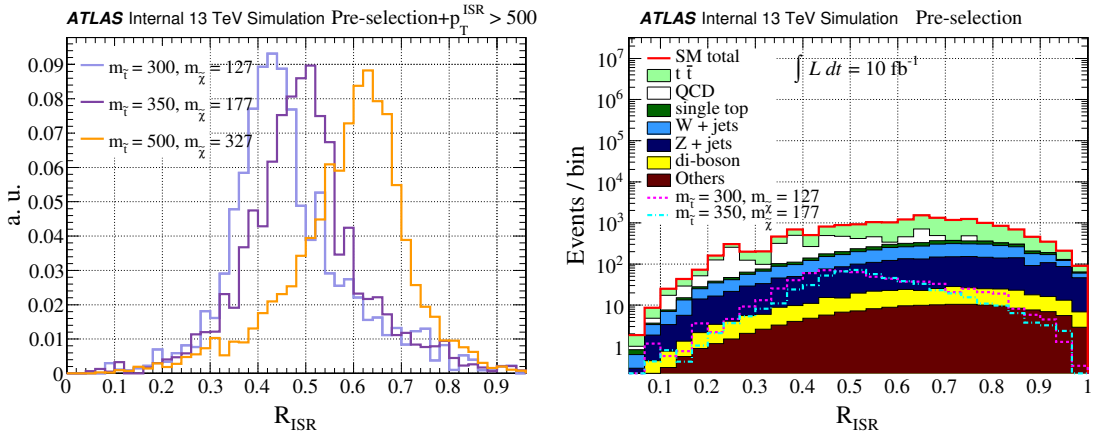

Figure A.2: Distributions of R_{ISR} for simulated samples, after the signal preselection is applied. *Left*: Three signal samples with differing $\tilde{t}_1$ and $\tilde{\chi}_1^0$ masses, with an additional $p_T^{\text{ISR}} > 500$ GeV requirement applied - in each case, the R_{ISR} distributions peak almost exactly at $m_{\tilde{t}_1}/m_{\tilde{\chi}_1^0}$. *Right*: Two signal samples are shown against the SM background predictions. The backgrounds are, in general, much flatter in R_{ISR} than the signal samples, with QCD multijet events generally falling at lower values.

A.1 Recursive Jigsaw Reconstruction for ISR-assisted E_T^{miss} Signals

For compressed SUSY scenarios, the presence of a high-p_T ISR system can *kick* the entire sparticle decay system into the transverse plane. The momentum received in this manner by massive invisible particles will appear as E_T^{miss} in the reconstructed event. However, the question becomes - how is the ISR system separated from the sparticle system, in order to find this missing energy?

Recursive Jigsaw Reconstruction (RJR) [99] is a general scheme for analysing high-energy collision events, both providing a way for deriving useful kinematic observables and implementing algorithms for overcoming combinatoric and kinematic ambiguities. While it has many use cases, both simple and complex, in analysing collisions, it is used here to reconstruct ISR-assisted sparticle decays. The general principle of the RJR technique is to impose a *decay tree* on the events, which can be incredibly specific and include all final state objects and intermediate states, or can be very generalised, capturing only the kinematic features relevant to the analysis. In this case, the decay tree is kept generic, containing only 3 final states, as shown in Figure A.3.

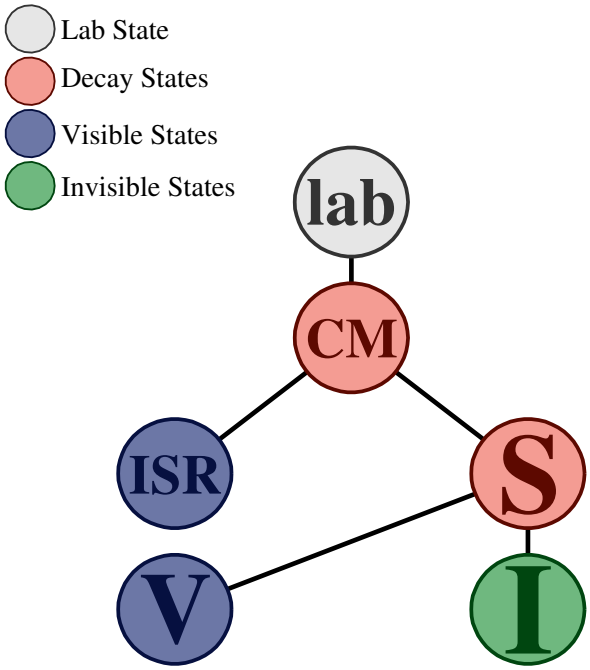

Figure A.3: Decay tree corresponding to the ISR-assisted E_T^{miss} signal analysis strategy. The sparticle system is denoted by S, and contains the non-ISR visible objects and the invisible states.

This decay tree constitutes a simplified picture of the events to be observed, with only three frames

(groups of particles, labelled V, I and ISR) identified in the final state. The sparticle frame (S) represents the system of massive particles produced in the event, which in this case is two stop quarks. This state decays to a collection of visible particles (the V frame) which we observse in the detector, and one or more invisible particles (I frame) which escape detection and are only observed indirectly through the missing transverse momentum. The ISR frame represents all objects in the event that do not come from the sparticle frame. The centre-of-mass frame, CM, is then the sum of all reconstructed and missing objects in these three frames. Each frame is identified with a four-vector sum of all particles, which can be evaluated in any of the reference frames implicitly defined in our decay tree. For example, the energy of the S state evaluated in the CM frame corresponds to the energy component of the four-vector associated with all the constituents of S evaluated in the rest frame of all the CM constituents.

Imposing this decay tree involves assigning objects to these frames and making either choices or guesses for any missing pieces of information (such as the mass of the object) required to assign a four-vector to each state. This is particularly relevant to the I frame, where the vectorial E_T^{miss} in each event is interpreted as the transverse momentum of I, evaluated in the lab frame. However this leaves two four-vector components unspecified; the mass of I, and its longitudinal momentum in the lab frame. While the mass can be set to zero, a not-necessarily correct assumption but rather a convenient one, the longitudinal momentum is tricky; rather than trying to guess it, it is deliberately ignored, restricting the entire event to the transverse plane.

To reconstruct the two visible states, assigning objects to either the V frame or the ISR frame, a *jigsaw* rule is applied, an interchangeable algorithmic procedure for resolving combinatoric ambiguity. Objects are grouped together based on closeness in phase-space in the lab transverse plane, before being assigned to a frame. This assignment is performed by simultaneously minimising the masses of the ISR and S frames, M_{ISR} and M_S, over all choices of the object assignment. The minimisation is done according to the following formula, choosing to maximise the ISR state and S state momentum in the CM frame:

$$M_{\text{CM}} = \sqrt{M_{\text{ISR}}^2 + \left(p_{\text{ISR}}^{\text{CM}}\right)^2} + \sqrt{M_{\text{S}}^2 + \left(p_{\text{S}}^{\text{CM}}\right)^2} \tag{A.2}$$

where $p_{\text{ISR}}^{\text{CM}} = p_{\text{S}}^{\text{CM}}$ are the momenta of the two systems, equal in magnitude and opposite in direction in the CM frame. Regardless of how objects are partitioned between ISR and V frames, the total mass M_{CM} is constant in each event - hence by maximising $p_{\text{ISR/S}}^{\text{CM}}$, the masses of the ISR and S frames are simultaneously minimised. This procedure is analogous to an exclusive two-jet reclustering of all objects in the event, treating the E_T^{miss} as another jet. For this reason, the ISR frame, which may actually be a complex system of multiple objects, is referred to throughout the bulk of this thesis as an ISR jet, wherever this technique has been applied.

With each reconstructed object in the event assigned to either the V frame or the ISR frame, and the mass of the I state specified, a full kinematic description of the event is obtained, in the transverse plane. Due to this deliberate ignorance to the longitudinal momentum in each event, the masses of the different states are actually the transverse masses, and are explicitly stated as such throughout the text.

Appendix B

QCD Multijet and All-hadronic $t\bar{t}$ Estimation

The background from the production of multijet events and all-hadronic $t\bar{t}$ events is estimated with the jet smearing method. This has been used previously by the $\tilde{b}_1$ pair production analysis [84], and several other ATLAS SUSY analyses [101]. While control regions and validation regions are defined here for the purpose of QCD background estimation, they are not fed into the simultaneous fit used to estimate the other background sources, as the QCD multijet contribution is small.

B.1 Methodology Overview

The jet smearing method is used to repeatedly smear the Lorentz vector of jets in well-measured data events with small E_T^{miss}, creating "pseudo-data" with potentially large E_T^{miss}. The main assumption of this method is that the QCD background is dominated by the mis-measurement of multiple jets, whereby the hadronisation of partons is not fully reconstructed by the detector, or the hadronisation produces real E_T^{miss} in the form of neutrinos. The different stages of the method are as follows:

1. Data events containing ≥ 4 jets, of which two are b-tagged, are used to select well measured "seed events" by defining the E_T^{miss} significance as:

$$E_T^{\text{miss}}\text{sig.} = \frac{E_T^{\text{miss}} - M}{\Sigma E_T} \tag{B.1}$$

   where M is a constant, found to be $M = 8$ GeV. The motivation for this was to remove any bias in the leading p_T jet distribution of the pseudo-data.

2. The momentum of jets in the seed events is then smeared - the Lorentz vector of each jet is multiplied by a random number derived from pre-determined jet response maps.

3. The operation described in step 2 is repeated 5000 times for each jet in the seed event to randomly generate configurations where the E_T^{miss} comes from multiple fluctuating jets.

B.2 Seed Event Selection

Seed events are selected from 2015 and early 2016 ATLAS data, and must have been recorded with any offline software jet trigger (`HLT_j*`). Each seed event is weighted based on the online leading p_T and which trigger fired. Events with leptons are vetoed, and at least four reconstructed jets (of which two are b-tagged) must be present in the event. A requirement on the E_T^{miss} significance (defined above) is applied based on the number of b-tagged jets in the event:

$$E_T^{\text{miss}}\text{sig.} < 0.3 + 0.1 \cdot n_{\text{b-tag}}. \tag{B.2}$$

Selected seed events are processed with the jet smearing method described above. Seed events with the smeared jets create the pseudo-data sets, which are compared to simulation in control and validation regions in order to estimate the multijet background in the signal regions.

B.3 Control and Validation Regions

To be selected for use in a control or validation region, pseudo-data must pass the following requirements. An E_T^{miss} trigger must be passed, with either $E_T^{\text{miss}} > 70$ GeV if from 2015 or $E_T^{\text{miss}} > 90$ if from 2016. Events with leptons are vetoed, and events must have at least 4 jets (2 b-tagged jets). The reconstructed E_T^{miss} must be greater than 200 GeV, and the signal region QCD-killing cut is inverted, $\min|\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})| < 0.4$. Table B.1 shows the requirements for the control and validation regions associated to SRA and SRB, and Tables B.2 and B.3 show the CRs and VRs for SRC and SRD. These regions are defined here, rather than in the main body of the text, as they are not used as input into the simultaneous fit; instead, the QCD background estimation is taken directly from MC in the SRs, with the CRs and VRs described here used for normalisation and validation of the method.

Region		CRQAB	VRQAB
	$\min\lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert$	< 0.1	[0.1,0.4]
	$m^0_{\text{jet},R=1.2}$	> 120 GeV	
	$m_T^{b,\text{min}}$	> 100 GeV	
	b-tagged jets	≥ 2	
	τ-veto	yes	
A	$m^0_{\text{jet},R=0.8}$	> 60 GeV	
	E_T^{miss}	> 300 GeV	
B	$m_T^{b,\text{max}}$	> 200 GeV	
	$\Delta R(b,b)$	> 1.2	

Table B.1: QCD control and validation region definitions for SRA and SRB. The signal preselection from Table 6.1 is applied in the VR.

Variable	CRC	VRC
$\min\lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert$	[0.05,0.1]	[0.1,0.2]
N^S_{btag}	≥ 1	
N^S_{jet}	≥ 5	
$p_{T,b}^{0,S}$	> 40 GeV	
m_T^S	-	> 300 GeV
$\Delta\phi_{\text{ISR},I}$	> 2.00	> 3.00
p_T^{ISR}	> 150 GeV	> 400 GeV
R_{ISR}	< 0.4	
$p_T^{4,S}$	> 50 GeV	
b-tagged jets	≥ 1	

Table B.2: QCD control and validation region definitions for SRC. The signal preselection from Table 6.1 is applied in the VR.

	CRQD	VRQD	VRQD-low	VRQD-high
$\min\lvert\Delta\phi(\text{jet}^{0,1}, E_T^{\text{miss}})\rvert$	< 0.1	[0.1,0.4]		
E_T^{miss}	> 250 GeV			
NJets	≥ 5			
b-tagged jets	≥ 2			
$\Delta R(b,b)$	> 0.8			
τ-veto	yes			
jet p_T^1	-		> 150 GeV	
jet p_T^3	-		> 100 GeV	> 80 GeV
jet p_T^4	-		> 60 GeV	
$m_T^{b,\min}$	-		> 250 GeV	> 350 GeV
$m_T^{b,\max}$	-		> 300 GeV	> 450 GeV
b-jet $p_T^0 + p_T^1$	-		> 300 GeV	> 400 GeV

Table B.3: QCD control and validation region definitions for SRD. The signal preselection from Table 6.1 is applied in the VRs.

Bibliography

[1] F. Halzen and A. Martin. *Quarks & Leptons: An Introductory Course in Modern Particle Physics*. John Wiley & Sons, 1984.

[2] R. Mann. *An Introduction to Particle Physics and the Standard Model*. Taylor and Francis Group, 2011.

[3] C. Patrignani et al. *Particle Data Group Handbook*. Particle Data Group, 2014.

[4] ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys. Lett. B*, 726(120), 2013.

[5] P. Z. Skands. Introduction to QCD. *arXiv:1207.2389*, 2014.

[6] LHCb Collaboration. Observation of the resonant character of the Z(4430)$^-$ state. *Phys. Rev. Lett.*, 112(222002), 2014.

[7] Belle Collaboration. Study of $e^+e^- \to \pi^+\pi^- J/\psi$ and observation of a charged charmoniumlike state at Belle. *Phys. Rev. Lett.*, 110(252002), 2013.

[8] BES-III Collaboration. Observation of a charged charmoniumlike structure in $e^+e^- \to \pi^+\pi^- J/\psi$ at $\sqrt{s} = 4.26$ GeV. *Phys. Rev. Lett.*, 110(252001), 2013.

[9] LHCb Collaboration. Observation of J/ψp resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays. *Phys. Rev. Lett.*, 115(072001), 2015.

[10] Particle Data Group. Electroweak Model and Constraints on New Physics. http://pdg.lbl.gov/2009/reviews/rpp2009-rev-standard-model.pdf.

[11] P. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13(508), 1964.

[12] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13(321), 1964.

[13] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716, 2012.

[14] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716, 2012.

[15] T2K Collaboration. Evidence of electron neutrino appearance in a muon neutrino beam. *Phys. Rev. D.*, 88(032002), 2013.

[16] Super-Kamiokande Collaboration. Evidence of oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81(1562), 1998.

[17] F. P. An et al. Observation of electron-antineutrino disappearance at Daya Bay. *Phys. Rev. Lett*, 108(171803), 2012.

[18] S. P. Martin. A Supersymmetry Primer. arXiv:hep-ph/9709356.

[19] I. J. R. Aitchison. Supersymmetry and the MSSM: an elementary introduction. arXiv:hep-ph/0505105.

[20] C. Csaki. The Minimal Supersymmetric Standard Model (MSSM). *Mod. Phys. Lett. A*, 11(599), 1996.

[21] G. Anderson and D. Castano. Naturalness and superpartner masses, or when to give up on weak scale supersymmetry. *Phys. Rev. D*, 52(1693), 1995.

[22] R. Catena and L. Covi. SUSY dark matter(s). *Eur. Phys. J. C.*, 74(2703), 2014.

[23] C. Borchensky, M. Kramer, et al. Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV. *Eur. Phys. J. C.*, 74(12)(3174), 2014.

[24] T. Cohen, T. Golling, et al. SUSY simplified models at 14, 33 and 100 TeV proton colliders. *JHEP*, 04(117), 2014.

[25] D. Alves et al. Simplified Models for LHC New Physics Searches. *J. Phys. G: Nucl. Part. Phys.*, 39(105005), 2012.

[26] K.J. de Vries MasterCode Project et al. The pMSSM10 after LHC Run 1. *Eur. Phys. J. C*, 75(422), 2015.

[27] E. A. Bagnaschi MasterCode Project et al. Supersymmetric Dark Matter after LHC Run 1. arXiv:1508.01173 [hep-ph].

[28] E. Bagnaschi MasterCode Project et al. Likelihood Analysis of Supersymmetric SU(5) GUTs. arXiv:1610.10084 [hep-ph].

[29] S. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D.*, 2(1285), 1970.

[30] L. Maiani. The GIM Mechanism: origin, predictions and recent uses. arXiv:1303.6154 [hep-ph].

[31] LHCb Collaboration. Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay. In *50th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, 2015.

[32] W. Altmannshofer and D. M. Straub. New Physics in $B \to K^*\mu\mu$? *Eur. Phys. J. C.*, 73(2646), 2013.

[33] CMS Collaboration. Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett. B.*, 753(424), 2016.

[34] The CMS and LHCb Collaborations. Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data. *Nature*, 522(68), 2015.

[35] LHCb Collaboration. Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays. arXiv:1703.05747 [hep-ex].

[36] A. Arbey, M. Battaglia, F. Mahmoudi, and D. Martinez Santos. Supersymmetry confronts $B_s^0 \to \mu^+\mu^-$: present and future status. *Phys. Rev. D.*, 87(035026), 2012.

[37] LHCb Collaboration. Letter of Intent for the LHCb Upgrade. cds.cern.ch/record/1333091. CERN-LHCC-2011-001.

[38] D. M. Straub. Overview of constraints on new physics in rare B decays. arXiv:hep-ph/1205.6094.

[39] O. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, K. Poole, and P. Proudlock. *LHC Design Report.* CERN, 2004.

[40] C. Lefevre and CERN. The CERN accelerator complex. http://cds.cern.ch/record/1260465, 2008.

[41] V. Papadimitriou. Luminosity Determination at the Tevatron. In *LHC Lumi Days: LHC Workshop on LHC Luminosity Calibration*, 2011.

[42] ATLAS Collaboration. Public luminosity results website. http://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2, 2015.

[43] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(S08003), 2008.

[44] M. Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. CERN / LHCC 2010-013, 2010.

[45] ATLAS Collaboration. Impact Parameter Resolution, Run 1 vs Run 2. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2015-007/.

[46] ATLAS Collaboration. Performance of Particle Identification with the ATLAS Transition Radiation Tracker. In *Proceedings of the DPF-2011 Conference*, 2011.

[47] F. Campabadal et al. Beam tests of ATLAS SCT silicon strip detector modules. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 538 - Issue 1-3:384–407, 2005.

[48] M. Bona, M. Castanheira, S. McMahon, E. Piccaro, and P. Ward. Lorentz Angle Analysis and Measurement for the ATLAS SCT. *ATLAS note*, ATL-COM-INDET-2011-114, 2011.

[49] C. Jacoboni, C. Canali, G. Ottaviani, and A. Alberighi Quaranta. A review of some charge transport properties of silicon. *Solid State Electonics*, 20:77, 1977.

[50] C. Canali, G. Majini, R. Minder, and G. Ottaviani. Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature. *IEEE Trans. Electron Devices*, ED-7:481, 1970.

[51] J. Becker, E. Fretwurst, and R. Klanner. Measurements of charge carrier mobilities and drift velocity saturation in bulk silicon of <111> and <100> crystal orientation at high electric fields. *Solid State Electronics*, 56:104–110, 2011.

[52] C. Alpigiani, M. Bona, and S. McMahon. Lorentz angle measurements for the ATLAS silicon strip. *ATLAS note*, ATL-COM-INDET-2013-034, 2013.

[53] C. Bobeth and other. $B_{s,d} \to l^+ l^-$ in the Standard Model with reduced theoretical uncertainty. *Phys. Rev. Lett.*, 112(101801), 2014.

[54] C. S. Huang, W. Liao, and Q. S. Yan. The Promising Process to Distinguish Supersymmetric Models with Large $\tan\beta$ from the Standard Model: $B \to X(s)\mu^+\mu^-$. *Phys. Rev. D*, 59(011701), 1999.

[55] S. R. Choudhury and N. Gaur. Dileptonic decay of B(s) mesons in SUSY models with large $\tan\beta$. *Phys. Lett. B*, 451(86-92), 1999.

[56] K. S. Babu and C. F. Kolda. Higgs mediated $B^0 \to \mu^+\mu^-$ in minimal supersymmetry. *Phys. Rev. Lett.*, 84(228-231), 2000.

[57] CMS and LHCb Collaborations. Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data. *Nature*, 522(68), 2015.

[58] ATLAS Collaboration. Study of the rare decays of B_s and B_0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector. *Eur. Phys. J. C*, 76(513), 2016.

[59] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05(026), 2006.

[60] T. Sjostrand, S. Mrenna, and P. Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178(852-867), 2008.

[61] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462(152), 2001.

[62] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth. A*, 506(250-303), 2003.

[63] ATLAS Collaboration. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data. *Eur. Phys. J. C*, 74(3130), 2014.

[64] ATLAS Collaboration. Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, 75(120), 2015.

[65] ATLAS Collaboration. Search for the decay $B_s^0 \to \mu^+\mu^-$ with the ATLAS detector. *Phys. Lett. B*, 713(387), 2012.

[66] A. Hoecker et al. TMVA 4, Toolkit for Multivariate Data Analysis with ROOT Users Guide. arXiv:physics/0703039.

[67] N. L. Johnson. Systems of frequency curves generated by methods of translation. *Biometrika*, 36(149-176), 1949.

[68] M. C. Jones and A. Pewsey. Sinh-arcsinh distributions. *Biometrika*, 96(761-780), 2009.

[69] ATLAS Collaboration. Measurement of the Υ(1S) Production Cross-Section in pp Collisions at $\sqrt{s} = 7$ TeV in ATLAS. *Phys. Lett. B*, 705(9), 2011.

[70] ATLAS Collaboration. Determination of the ratio of b-quark fragmentation fractions f_s/f_d in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev. Lett.*, 115(262001), 2015.

[71] J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans. R. Soc. London. A*, 236(333-380), 1937.

[72] G. Cowan et al. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71(1554), 2011.

[73] A. L. Read. Presentation of search results: The CL_s technique. *J. Phys. G*, 28(2693-2704), 2002.

[74] S. Weinberg. Implications of Dynamical Symmetry Breaking. *Phys. Rev. D*, 13(974), 1976.

[75] S. Dimopoulos, S. Raby, and F. Wilczek. Supersymmetry and the scale of unification. *Phys. Rev. D*, 24(1681), 1981.

[76] P. Fayet. Supersymmetry and weak, electromagnetic and strong interactions. *Phys. Let. B*, 64(159), 1976.

[77] W. Beenakker et al. Supersymmetric top and bottom squark production at hadron colliders. *JHEP*, 08(098), 2010.

[78] W. Beenakker et al. Squark and gluino hadroproduction. *Int. J. Mod. Phys. A*, 26(2637), 2011.

[79] J. Ellis et al. Supersymmetric relics from the Big Bang. *Nucl. Phys. B*, 238(453), 1984.

[80] T. Lin, E. W. Kolb, and L. T. Wang. Probing dark matter couplings to top and bottom quarks at the LHC. *Phys. Rev. D*, 88(063510), 2013.

[81] D. Abercrombie et al. Dark Matter benchmark models for early LHC Run-2 searches: report of the ATLAS/CMS dark matter forum. arXiv:1507.00966 [hep-ex].

[82] ATLAS Collaboration. Search for a scalar partner of the top quark in the Jets+E_T^{miss} Final State at $\sqrt{s} = 13$ TeV with the ATLAS detector. ICHEP 2016, cds.cern.ch/record/2206250.

[83] ATLAS Collaboration. Search for direct pair production of the top squark in all-hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 09(015), 2014.

[84] ATLAS Collaboration. Search for bottom squark pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 76(547), 2016.

[85] CERN. LHC performance reaches new highs.

[86] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross-sections, and their matching to parton shower simulations. *JHEP*, 07(079), 2014.

[87] M. Kramer et al. Supersymmetry production cross sections in p collisions at $\sqrt{s} = 7$ TeV. arxiv:1206.2892 [hep-ph], 2012.

[88] T. Gleisberg et al. Event generation with SHERPA 1.1. *JHEP*, 02(007), 2009.

[89] S. Alioli et al. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06(043), 2010.

[90] ATLAS Collaboration. Vertex reconstruction performance of the ATLAS detector at $\sqrt{s} = 13$ TeV, 2015.

[91] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k(t) jet clustering algorithm. *JHEP*, 04(063), 2008.

[92] M. Cacciari and G. P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659(119), 2008.

[93] ATLAS Collaboration. Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector. ATLAS-CONF-2015-029, 2015.

[94] on behalf of the ATLAS Collaboration A. Calandri. Flavour tagging algorithms and performance at the ATLAS experiment. In *4th Conference on Large Hadron Collider Physics 2016 (LHCP 2016)*, 2016.

[95] ATLAS Collaboration. Optimisation of the ATLAS b-tagging performance for the 2016 LHC run, 2016.

[96] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV. arXiv:1603.05598 [hep-ex], 2016.

[97] H. An and L. T. Wang. Opening up the compressed region of top squark searches at 13 TeV LHC. *Phys. Rev. Lett.*, 115(181602), 2015.

[98] S. Macaluso et al. Revealing compressed stops using high-momentum recoils. *JHEP*, 03(151), 2016.

[99] P. Jackson, C. Rogan, and M. Santoni. Sparticles in motion - getting to the line in compressed scenarios with the recursive jigsaw reconstruction. arXiv:1607.08307 [hep-ph], 2016.

[100] M. Baak et al. HistFitter software framework for statistical data analysis. *Eur. Phys. J. C*, 75(153), 2015.

[101] ATLAS Collaboration. Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using 4.7 fb^{-1} of $\sqrt{s} = 7$ TeV proton-proton collision data. *Phys. Rev. D*, 87(012008), 2013.

[102] S. Ask, M. A. Parker, T. Sandoval, M. E. Shea, and W. J. Stirling. Using gamma+jets production to calibrate the Standard Model Z(nunu)+jets Background to New Physics Processes at the LHC. *JHEP*, 10(058), 2011.

[103] ATLAS Collaboration. Jet energy resolution and reconstruction efficiencies from in-situ rechniques with the ATLAS detector. ATLAS-CONF-2010-054, 2010.

[104] ATLAS Collaboration. Jet energy scale and its systematic uncertainty in proton-proton collisions. ATLAS-CONF-2013-004, 2013.

[105] ATLAS Collaboration. Summary of recent SUSY search results. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SUSY/.

[106] ATLAS Collaboration. Summary of the searches for squarks and gluinos using $sqrts = 8$ TeV pp collisions with the ATLAS experiment at the LHC. *JHEP*, 10(054), 2015.

[107] B. Nachman. Less is more when gluinos mediate. arXiv:1505.00994 [hep-ph], 2015.

[108] R. D. Cousins, J. T. Linnemann, and J. Tucker. Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hyopthesis for a Poisson process. *Nucl. Inst. Meth. A*, 595(480-501), 2008.

[109] K. Cranmer. Statistical challenges for searches for new physics at the LHC. arXiv:physics/0511028.

[110] ATLAS Collaboration. Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Phys. Rev. D*, D94(032005), 2016.

[111] ATLAS Collaboration. Search for dark matter produced in association with a Higgs boson decaying to two bottom quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. D*, 93(072007), 2016.

[112] ATLAS Collaboration. Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 75(299), 2015.

[113] ATLAS Collaboration. Search for dark matter in events with heavy quarks and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 75(92), 2015.

[114] Seungwon Baek, P. Ko, and Jinmian Li. Minimal renormalizable simplified dark matter model with a pseudoscalar mediator. arXiv:1701.04131.

[115] B. Bhattacherjee, A. Choudhury, K. Ghosh, and S. Poddar. Compressed SUSY at 14 TeV LHC. *Phys. Rev. D*, 89(037702), 2014.

[116] A. Nelson, P. Tanedo, and D. Whiteson. Limiting SUSY compressed spectra scenarios. *Phys. Rev. D*, 93(115029), 2016.

[117] P. Baldi, P. Sadowski, and D. Whiteson. Searching for exotic particles in high-energy physics with deep learning. arXiv:1402.4735 [hep-ph], 2014.

[118] A. Buckley, A. Shilton, and M. J. White. Fast supersymmetry phenomenology at the Large Hadron Collider using machine learning techniques. arXiv:1106.4613 [hep-ph], 2011.

[119] S. Caron, J. S. Kim, K. Rolbiecki, R. R. de Austri, and B. Stienen. The BSM-AI project: SUSY-AI - Generalizing LHC limits on supersymmetry with machine learning. arXiv:1605.02797 [hep-ph], 2016.

[120] G. Cowan. Topics in statistical data analysis for high-energy physics. arXiv:1012.3589 [physics.data-an].

[121] D. M. Hawkins. The problem of overfitting. *J. Chem. Inf. Comput. Sci.*, 44(1), 2004.

[122] I. V. Tetko, D. J. Livingstone, and A. I. Luik. Neural network studies 1: Comparison of overfitting and overtraining. *J. Chem. Inf. Comput. Sci.*, 35(5), 1995.

[123] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 1958.

[124] A. A. Rusu et al. Progressive Neural Networks. arXiv:1606.04671 [cs.LG], 2016.

[125] J. Schmidhuber. Deep learning in neural networks: an overview. *Neural Networks*, 61(85-117), 2015.

[126] L. Rokach and O. Maimon. *Data mining with decision trees: theory and applications*. World Scientific Pub Co Inc, 2008.

[127] J. H. Friedman. Stochastic gradient boosting. *Comp. Stats. D. An.*, 38(4), 2002.

[128] L. Breiman. Bagging predictors. *Machine Learning*, 24, 1996.

[129] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Chem. Inf. Comput. Sci.*, 55(119-139), 1997.

[130] S. Arlot. A survey of cross-validation procedures for model selection. *Statistical Surveys*, 4(40-79), 2010.

[131] ATLAS Collaboration. Search for squarks and gluinos in final states with jets and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector. https://cds.cern.ch/record/2258145, 2017.

[132] J. Fan, M. Reece, and J. T. Ruderman. Stealth Supersymmetry. *JHEP*, 11(012), 2011.

[133] S. S. AbduSalam, B. C. Allanach, F. Quevedo, F. Feroz, and M. Hobson. Fitting the Phenomenological MSSM. *Phys. Rev. D*, 81(095012), 2010.

[134] ATLAS Collaboration. Summary of ATLAS Constraints in the pMSSM. *JHEP*, 10(134), 2015.

[135] ATLAS Collaboration. Search for resonances in diphoton events at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 09(001), 2016.